



Aalto-yliopisto
Sähkötekniikan
korkeakoulu

Vili Haukkovaara

**Suomen ajallisen paikkaontologian hyödyntäminen arkistolai-
toksen paikannimitiedon rikastamisessa**

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi
diplomi-insinöörin tutkintoa varten.

Espoossa 31.7.2015

Valvoja: Professori Eila Järvenpää

Ohjaaja: Matias Frosterus

Tekijä Vili Haukkovaara

Työn nimi Suomen ajallisen paikkaontologian hyödyntäminen arkistolaitoksen paikannimitiedon rikastamisessa

Laitos Sähkötekniikan korkeakoulu

Professori Työpsykologia ja johtaminen

Professuurikoodi TU-53

Työn valvoja Eila Järvenpää

Työn ohjaaja(t) Matias Frosterus

Päivämäärä 31.7.2015

Sivumäärä 60+5

Kieli Suomi

Tämän työn tavoitteena on selvittää miten hyvin Suomen ajallinen paikkaontologia soveltuu arkistolaitoksen paikannimitiedon rikastamiseen. Työssä käydään läpi arkistokuvailun ja arkistolaitoksen aineistojen metatietojen hallinnointijärjestelmien nykytilaa sekä asiakkaiden odotuksia muistiorganisaatioita kohtaan. Työssä kuvataan myös tavoitetila, jossa arkistokuvailu on uudistunut ja käytössä on uusi arkistojen metatietojen hallinnointiväline. Tavoitetila luo muutospaineen myös paikannimitiedon rikastamiselle.

Työssä kuvataan teknologia ja Suomen ajallinen paikkaontologia, jonka avulla paikannimitietoa voitaisiin rikastaa. Ratkaisussa arkistoaineistojen nykyisistä metatiedoista poimitaan paikannimiä ja näille pyritään löytämään vastineet Suomen ajallisesta paikkaontologiasta. Ratkaisun toteutuksen yhteydessä syntyvän datan avulla voidaan osoittaa, että Suomen ajallinen paikkaontologia on hyvä ratkaisu arkistolaitokselle, mutta ei kuitenkaan täysin riittävä tavoitetilan kannalta.

Avainsanat arkistot, avoin linkitetty data, SAPO, ontologiat, paikkatieto, arkistokuvailu

Author Vili Haukkovaara

Title of thesis Enriching archival spatial metadata with Finnish spatio-temporal ontology

Department School of electrical engineering

Professorship Work Psychology and Leadership**Code of professorship** TU-53

Thesis supervisor Eila Järvenpää

Thesis advisor(s) / Thesis examiner(s) Matias Frosterus

Date 31.7.2015**Number of pages** 60+5**Language** Finnish

The purpose of this thesis is to study if Finnish spatio-temporal ontology is good source of spatial information enrichment for the metadata which the National Archives service of Finland manages. The study describes the present state of the archival description in Finland and the systems which are used. Also the needs of the customers of the memory organizations are discussed.

In the future the National Archives of Finland will have a new archival system which datamodel is based on the new archival description model. This creates a need to enrich current archival metadata including spatial metadata.

The study describes the technology used and the Finnish spatio-temporal ontology. In the solution spatial metadata from the archival metadata is extracted and found place names are being mapped to Finnish spatio-temporal ontology. The data which is produced during mapping provide us the answer that the Finnish spatio-temporal ontology is good source for enriching current metadata but it is not adequate.

Keywords archives, open linked data, SAPO, ontologies, geographical data, archival description

Alkusanat

Opiskeluni ovat venyneet ja diplomi-insinööriksi valmistuminen on ollut osaltaan riippuvainen diplomityön puuttumisesta opinnoistani. Nyt, perheenisäyksen lähestyessä, oli sopiva aika viedä yksi elämänvaihe loppuun toisen alkaessa. Työn aiheeksi valikoitui opinnoistani vahvasti mieleen jäänyt semanttiseen webiin ja ontologioihin liittyvä, mutta myös nykyisen työnantajani Kansallisarkiston, intresseissä oleva aihe.

Haluan kiittää erityisesti kollegaani Antti Laukkosta, sillä hän on mahdollistanut omalla panoksellaan työni etenemisen. Haluan myös kiittää valvojaani Eila Järvenpäää kiitettävästä ohjauksesta ja antoisista keskusteluista työni etenemisen ohella. Myös valvojaa-ni Matias Frosterusta haluan kiittää, sillä hän oli suurena apuna työni alkuvaiheessa ja aiheen muotoutuessa. Omat kiitoksensa ansaitsee myös yksi inspiraation lähteistäni, arkistoneuvos Jaana Kilkki sekä oma esimieheni Päivi Happonen, joka on mahdollistanut joustavan osallistumiseni valvojani kanssa työskentelyyn. Viimeisenä kiitoksen koh-teenä, muttei missään nimessä vähäisimpänä haluan kiittää vaimoani Katri Haukkovaa-raa, joka on jaksanut kuunnella valitustani ja tuskasteluani työni edetessä.

Espoo 31.7.2015

Vili Haukkovaara

Sisällysluettelo

1. Johdanto	5
1.1. Nykyisen toimintaympäristön haasteet arkistoissa sekä tavoiteltu muutos.....	5
1.2. Tutkimuksen rajaus	7
1.3. Työn rakenne	8
2. Arkistolaitoksen järjestelmien ja asiakkaiden tarpeiden nykytila.....	10
2.1. Kansainväliset standardit arkistokuvailussa	10
2.2. Arkistojen asiakkaiden tarpeet paikkoja koskevasta tiedosta.....	11
2.3. Arkistolaitoksen nykyjärjestelmät ja niiden sisältämä paikkatieto	16
2.4. Yhteenveto nykytilanteesta ja siinä havaituista ongelmista	19
3. Arkistolaitoksen järjestelmien ja palveluiden tavoitetilä.....	20
3.1. Ohjaavat rakenteet ja sidosryhmät.....	20
3.1.1. KDK – Kansallinen digitaalinen kirjasto	21
3.1.2. Finna.....	24
3.1.3. Finto	26
3.2. AHAA tietorakenne	26
3.3. Paikkaan liittyvien metatietojen hallinta AHAA-järjestelmässä.....	29
4. Teknologia ja sen hyödynnettävyys.....	31
4.1. Teknologiat ratkaisun taustalla.....	31
4.1.1. WWW – World Wide Web	31
4.1.2. Semanttinen web	32
4.1.3. Ontologiat.....	34
4.1.4. Avoin linkitetty data.....	35
4.2. Ontologioiden hyödyt.....	37
4.2.1. Ontologioiden hyödynnettävyys tiedon hakemisessa	38
4.2.2. Ontologisoidun paikkatiedon hyödyntäminen hakemisessa	40
4.3. SAPO.....	40

5.	Ratkaisun vaatimukset ja ratkaisun kuvaus	42
5.1.	Valittu aineisto	42
5.2.	Arviointimenetelmä	43
5.3.	Vaatimukset toteutukselle	43
5.4.	Toteutuksen kuvaus	46
6.	Tulokset	48
6.1.	Arkistonmuodostaja -kuvailukohteista löydetty paikannimistö ja SAPO	49
6.2.	Arkistoyksikkö kuvailukohteista löydetty paikannimistö ja SAPO	51
6.3.	Tulosten arviointi	54
7.	Yhteenveto ja pohdintaa	58
7.1.	Yhteenveto.....	58
7.2.	Pohdinta.....	59
8.	Lähdeluettelo	61

Lyhenteet

AHAA-hanke	Arkistojen hakemistopalveluhanke
Arkistolaitos	Arkistolaitoksen muodostavat Kansallisarkisto ja sen alaiset seitsemän maakunta-arkistoa.
Arkistoyksikkö	Arkistoyksikkö on sarjaa pienempi arkiston (asiakirjakokonaisuuden) osa. (Arkistolaitos, ei pvm)
	Description
Finna	Finna on tiedonhakupalvelu, joka kokoaa yhteen aineistoja arkistoista, kirjastoista ja museoista.
Finto	Suomalainen sanasto- ja ontologiapalvelu, joka mahdollistaa sanastojen julkaisun ja selailun. (Finto, 2015)
	for Corporate Bodies, Persons and Families
ICA	International Council on Archives
ISAAR (CPF)	International Standard Archival Authority Record
ISAD (G)	General International Standard Archival
Kansallisarkisto	Kansallisarkisto johtaa arkistolaitoksen toimintaa. Se ohjaa valtion keskushallinnon arkistointia sekä toimii valtakunnallisena keskusvirastona ja toimialansa tutkimus- ja kehittämisskeskuksena.
KOKO	KOKO on kokoelma suomalaisia ydinontologioita, jotka on liitetty peilauksilla toisiinsa. (Finto, ei pvm)
Metatieto	Metatieto on tietoa tiedosta. Se on kuvailevaa ja määrittelevää tietoa, jostakin kohteesta. Esimerkiksi ”auton väri” voi olla ”auton” metatieto.
Muistiorganisaatiot	Yleisesti tarkoitetaan kirjastoja, arkistoja ja museoita yhdessä.
OKM	Opetus- ja kulttuuriministeriö
Paikkatieto	Paikkatiedolla (geographic information, spatial data) tarkoitetaan kaikkea tietoa, joka sisältää viittauksen tiettyyn paikkaan tai maantieteelliseen alueeseen (Maanmittauslaitos, 2015).
SAPO	Suomen ajallinen paikkaontologia. Kuvaa historiallisten ja nykyisin olemassa olevien suomalaisten paikkojen ajallisen ja spatiaalisen ulottuvuuden. (Finto, 2014)

SeCo	The Semantic Computing Research Group
WGS84	WGS 84 (World Geodetic System) on Yhdysvaltain puolustusministeriön määrittelemä ja ylläpitämä koordinaattijärjestelmä ja siihen liittyvä geoidimalli. (Wikipedia, 2014)
YSO	Yleinen suomalainen ontologia. YSO pohjautuu yleiseen suomalaiseen asiasanastoon (YSA) sekä yleiseen ruotsinkieliseen tesaurukseen (Allärs). (Finto, ei pvm)

1. Johdanto

Tässä työssä tarkastellaan muutosta arkistoalalla ja muutoksen myötä syntynyttä muutospainetta. Muutospaineen alla on kyettävä löytämään uusia tapoja toimia ja tässä työssä on kuvattu arkistolaitoksen tavoitetilaa ja siihen pääsemiseksi on kehitetty ratkaisumalli, jonka avulla pyritään selvittämään miten SAPO (Suomen ajallinen paikkaontologia) soveltuu arkistolaitoksen paikannimitiedon rikastamiseen nykyisessä muodossaan.

1.1. Nykyisen toimintaympäristön haasteet arkistoissa sekä tavoiteltu muutos

Arkistojen kansainvälisesti standardoidut kuvailusäännöt ja käsitemallit ovat peräisin 1990-luvulta ajalta ennen internetin globalisoitumista. Viimeisten vuosien aikana arkistoinstituutiot ovat huomanneet, että vanhat käytännöt ja toimintatavat eivät toimi nykyisessä muuttuneessa toimintaympäristössä. Toimintaympäristön muutoksena pidetään arkistoissa yleisesti muutosta, jossa palvelut siirtyvät entistä enemmän verkkoon ajasta ja paikasta riippumattomiksi. Tämä taas on luonut muutospaineen arkistoinstituutioissa, sillä palvelu on perustunut henkilökohtaiseen ihmiseltä ihmiselle annettuun palveluun. Tätä on ollut mahdollista ylläpitää pitkään, sillä suuri osa aineistoihin liittyvästä tiedosta on ollut perinteisesti kokeneiden asiakaspalvelijoiden päässä, ennemminkin kuin tallennettuna esimerkiksi tietorakennemuotoon.

Kun palvelu siirtyy verkkoon, kasvavat odotukset myös sähköisessä muodossa olevaa metatietoa kohtaan. Käyttäjille on kyettävä tarjoamaan laadukkaita hakujärjestelmiä, joiden avulla voidaan hakea aineistoja. Käyttäjälle tulee voida esimerkiksi ehdottaa vaihtoehtoisia synonyymejä hakutermeistä, kuten tarjoamalla ”hilla” –hakusanalla haettaessa myös ”lakka-aiheet” hakutulokset. Haut kohdistuvat aineistojen metatietoihin, jolloin myös metatietojen tulee olla laadukkaita. Tästä johtopäätöksenä voidaan todeta, että asiakkaille verkossa tarjottujen palveluiden laatu ei ole riippuvaista ainoastaan itse palvelusta ja sen toiminnallisuuksista vaan myös siitä datasta, jota palvelussa on käytettävissä. Arkistojen osalta käytännössä ainoina metatietoina aineistoista voidaan pitää aineistojen kuvailu- ja luettelointitietoja.

Arkistolaitoksen ensimmäinen yhteinen tietojärjestelmä, jossa arkistoaineistojen kuvailu- ja luettelointitietoja on tallennettu ja käsitelty sähköisesti on 1990-luvun lopulta. Viime vuosien muutospaine kansallisesti ja kansainvälisesti arkistokuvailun uudistami-

seen on aiheuttanut muutospaineen myös järjestelmän ja etenkin arkistokuvailussa käytettävän tietomallin kehittämisessä. Vanha järjestelmä ei kuitenkaan toimi kehitysalustana uuden tietomallin osalta, vaan vuoden 2012 alusta käynnistettiin useiden suomalaisten arkistojen yhteinen AHAA-hanke (Arkistojen hakemistopalvelu-hanke,), jonka tavoitteena on toteuttaa uusi arkistojen yhteinen hakemistopalvelu, jossa hallinnoidaan erilaisten aineistojen kuvailutietoja (Arkistolaitos, 2012). Tavoitteena on kyetä hallinnoimaan aineistojen metatietoja riippumatta siitä, onko aineisto digitaalisessa tai analogisessa muodossa (paperi, mikrofilmi yms.).

AHAA-hanke jatkuu vuodelle 2016, jolloin järjestelmä otetaan käyttöön. Tätä ennen kuitenkin kunkin organisaation on migroitava eli siirrettävä nykyisten järjestelmien sisältämät metatiedot uuteen rakennettavaan järjestelmään. Migraatioiden yhteydessä arkistolaitoksessa tullaan suorittamaan metatiedon rikastamista erilaisin keinoin, tavoitteena parantaa metatiedon laatua erilaisia hakupalveluita varten. Yksi näistä keinoista on luoda yhteys kansallisesti ylläpidettyihin avoimen linkitetyn datan lähteisiin. Kansallisesti ylläpidettyjä, eli viranomaisen rahoittamia avoimen linkitetyn datan lähteitä ovat mm. Finto-projektissa (<http://finto.fi/fi/>) tuotettavat yhteiset ontologiapalvelut.

Arkistolaitos on valinnut asiasanoituksen ja aiheen määrittämiseen kohdeontologioiksi KOKOn (kokoelma suomalaisia ydinontologioita), joka sisältää YSO:n (Yleinen suomalainen ontologia), joka taas pohjautuu nykyjärjestelmissä käytettyyn YSA-sanastoon (Yleinen suomalainen asiasanasto) (Finto, ei pvm). Lisäksi arkistolaitoksen tavoitteena on saada luotua linkit ontologioihin, joissa ylläpidetään paikannimistöä ja paikkoihin liittyviä metatietoja. Paikannimet ovat haasteellisia siitä syystä, että ne ovat nykyjärjestelmissä vain tekstinä ja näistä ei ole saatavissa automaattisesti helposti lisätietoja. Toisen haasteen synnyttää Finton nykyisin ylläpitämät paikannimiä sisältävät ontologiat. Nykyisin Finto tarjoaa suomalaisen aikapaikka -ontologian SAPOn (Suomen ajallinen paikkaontologia), joka sisältää Suomen kuntahistoria ajalta Suomen kunnat ja näiden ajalliset kattavuudet. Lisäksi SAPO sisältää jonkin verran koordinaattitietoa.

Syy, miksi arkistolaitos haluaa käyttää vain kansallisesti ylläpidettyjä linkitetyn avoimen datan lähteitä, johtuu toimintaympäristöstä. Arkistolaitoksella on huomattavasti suuremmat vaikutusmahdollisuudet kansallisten ontologiapalveluiden kehitykseen ja ylläpitoon kuin esimerkiksi kansainvälisiin lähteisiin. Tarve kansallisiin ontologioihin vaikuttamiselle syntyy siitä, että arkistolaitos tulee uuden kuvailujärjestelmän valmistuttua käyttämään päivittäisessä kuvailutyössään suoraan avointa linkitettyä dataa. Ta-

voitteena on myös välttää kaikkea tiedon kopiointia ulkoisista ontologioista ja edistää ennemminkin sitä, että mahdollisimman useat tahot käyttäisivät samoja ontologioita. Tämä johtuu siitä, että arkistolaitoksen keskeisimmät sidosryhmät; muut arkistot, kirjastot ja museot on vahvasti ohjattu siirtymään käyttämään Finna-palvelua(<https://www.finna.fi/>). Finna on haku- ja palveluportaali, jossa saman käyttöliittymän kautta pääsee käsiksi kirjastojen, arkistojen ja museoiden aineistoihin ja näiden metatietoihin. Tällöin on järkevämpää muodostaa Finnan ja Finton välille vahvempi yhteys, jossa vain Finna hyödyntää Finton ontologioiden laajempia metatietoja, sen sijaan, että jokainen kirjasto, arkisto ja museo tekisi tämän erikseen.

1.2. Tutkimuksen rajaus

Tämä työ keskittyy selvittämään sen, miten hyvin arkistolaitoksen kuvailujärjestelmän sisältämä paikannimitieto saataisiin linkitettyä Finton tarjoamaan paikannimistöä sisältämään ontologian termeihin. Arviointi sisältää selvityksen siitä, miten hyvin paikannimille voidaan identifioida kohteet SAPOsta eli suomalaisesta aika-paikkaontologiasta. Arviointi sisältää analyysin siitä, kuinka usein prosentuaalisesti paikannimelle voidaan löytää yksiselitteinen vastine SAPOsta ja kuin monelle prosentille SAPOsta löytyy useampi vastine. Työ myös sivuaa arkistojen luettelointikäytäntöjä sekä sitä, miten paljon paikannimistöä on ylipäättään käytetty arkistokuvailussa. Työ pyrkii vastamaan tutkimuskysymykseen: ”Miten hyvin Suomen ajallinen paikkaontologia soveltuu arkistolaitoksen paikannimitiedon rikastamiseen?”.

Työssä rajaudutaan selvittämään paikkatietoa vain arkistolaitoksen Vakka-arkistotietokannan (<http://www.narc.fi:8080/VakkaWWW/EtuSivu.action>) osalta. Vastavaa ratkaisua voidaan kuitenkin soveltaa eri arkistotietokannoille ja se onkin perusteltua, sillä eri arkistotietokannoissa on käytetty eri luettelointiperiaatteita metatiedon tuottamisessa ja paikannimitiedon tuottamisessa.

Työssä rajaudutaan vain kansallisiin ja todennäköisimmin ylläpidettyihin paikkatietoontologioihin. Kansainvälisiä paikkatieto-ontologioita ei käsitellä työssä, sillä tarkoituksena on pitäytyä julkisin varoin ylläpidettyjen palveluiden hyödyntämiseen, jolla voidaan jossain määrin olettaa olevan pidempi elinkaari. Eri ontologioiden soveltuvuuden arviointia arkistojen näkökulmasta ei myöskään tehdä, sillä hyödynnettävät ontologiat ovat työn näkökulmasta jo ennalta määriteltäviä. Rajaus SAPOn käytön osalta on tehty toimintaympäristön perusteella. Koska arkistoinstituutiot ovat suurten aineistomassojen

takia pakotettu kuvailemaan aineistojaan kokonaisuuksien tasolla, on perusteltua, että tuotettua tietoa verrataan paikkatieto-ontologiaa, jossa tarkkuustaso ei ole liian suuri. SAPO soveltuu tähän hyvin, sillä se kattaa tiedot tarkimmillaan vain Suomen kunnista.

Työssä on käytetty konstruktiiivisen tutkimusmenetelmän mukaista lähestymistapaa. Konstruktiiivinen menetelmä tai lähestymistapa tarkoittaa ongelman ratkaisua kehitettyjen mallien, suunnitelmien yms. pohjalta (Kasanen, et al., 1993). Konstruktiiivinen menetelmä koostuu usein ongelman kuvauksesta, teoriaosuudesta, ongelman ratkaisun kuvaamisesta reaalimaailmassa sekä yhteenvedosta aiheeseen liittyen (Kasanen, et al., 1993).

1.3. Työn rakenne

Työn jäsentyy neljään osioon. Ensimmäisessä osiossa, luvussa 2, esitellään työhön liittyvää taustatietoa arkistokuvailusta ja arkistoasiakkaiden tarpeista. Lisäksi esitellään arkistolaitoksen nykyisiä arkistoaineistojen metatietojen hallinnointijärjestelmiä. Osio keskittyy nykytilan kuvaamiseen arkistolaitoksen hallinnoiman paikannimitiedon osalta. Luvun 2 lopussa kuvataan vielä arkistolaitoksen järjestelmien sekä niiden sisältävän metatiedon nykytila ja osoitetaan, että nykytilanteessa arkistolaitoksella ei ole edellytyksiä tyydyttää asiakkaidensa tarpeita parhaalla mahdollisella tavalla. Tämä osio toimii osaltaan konstruktiiivisen menetelmän ongelman kuvausosiona.

Toisessa osiossa, luvussa 3, käsitellään arkistolaitoksen tavoitteita. Arkistolaitoksen tavoitetilaan vaikuttavat muuttuva arkistokuvailu, muuttuva toimintaympäristö ja yhteiskäyttöiset sähköiset palvelut. Tavoitetilan muodostumiseen vaikuttavat vahvasti myös ulkopuoliset sidosryhmät. Edellä mainittujen asioiden ymmärtäminen auttaa lukijaa hahmottamaan miksi työssä käytetty ratkaisu on valittu. Tavoitetilassa asiakkaiden tarpeet voidaan tyydyttää nykyistä paremmin. Samalla osoitetaan myös se, että tavoitetilaan päästäkseen on kyettävä löytämään ratkaisu mm. paikannimitiedon rikastamiseen. Tämä osio toimii tietyllä tapaa teoriaosuutena ja toisaalta reaalimaailman kuvauksena konstruktiiivisessä menetelmässä.

Kolmannessa osiossa, luvussa 4, tarkastellaan aihealueeseen liittyvää teknologiaa ja sen hyödyntämismahdollisuuksia. Luvussa esitellään siten tausta ja mahdollisuudet, jotka ovat vaikuttaneet edellisessä luvussa esitettyyn tavoitetilaan. Osio toimii myös teoriaosuutena, jossa tarkastellaan avoimen linkitetyn datan taustalla olevaa teknologiaa

sekä ontologioiden käytöllä saavutettavia hyötyjä yleisellä tasolla. Lisäksi tarkemmin osoitetaan, että ontologioita voidaan hyödyntää, kun kohteita haetaan paikkatiedon perusteella. Teoriaosuudessa esitellään myös ratkaisussa käytettävä Suomen ajallinen paikkaontologia, jolla keskeinen rooli tavoitetilassa ja työn tutkimuskysymykseen vastaamisessa. Tämäkin osio toimii siten konstruktiivisen menetelmän teoriaosuutena.

Neljännessä osiossa, luvuissa 5, 6 ja 7, kuvataan ratkaisu, ratkaisun analysointi ja tuotetaan yhteenveto työstä. Luvussa 5 on myös perusteltu tarkemmin tutkimusaineiston valinta ja esitelty ratkaisun arviointimenetelmä. Ratkaisussa viitataan edellisissä osissa kuvattuihin teknologioihin ja esitellään miten näitä käytännössä on hyödynnetty. Ratkaisuehdotusten analysointi ja yhteenveto peilautuvat ensisijaisesti kuvattuun nykytilaan ja siinä esiin nostettuihin ongelmakohtiin, mutta myös ensimmäisessä osiossa esiin tuotuihin arkistoasiakkaiden tarpeisiin. Lisäksi arvioidaan niiltä osin kuin on mahdollista, miten toteutettu ratkaisu palvelee asetettua tavoitetilaa. Tämä osio sisältää konstruktiivisen menetelmän ratkaisun kuvaksen sekä yhteenvedon.

2. Arkistolaitoksen järjestelmien ja asiakkaiden tarpeiden nykytila

Tässä luvussa esitellään arkistojen käyttämiä metatietoja arkistokuvailussa nykytilassa sekä arkistojen asiakkaiden tarpeita kohdistuen paikkatietoon. Lisäksi luvussa esitellään arkistolaitoksen nykyjärjestelmiä ja niistä löytyviä metatietoja. Luvussa osoitetaan nykytilanteen ongelmat, jotka muodostuvat siitä, että asiakkaiden tarpeita ei kyetä tyydyttämään nykyisillä käytännöillä sekä järjestelmillä.

2.1. Kansainväliset standardit arkistokuvailussa

Arkistojen kansainvälisiä kuvailu- ja luettelointisääntöjä hallinnoi kansainvälinen järjestö ICA (The International Council on Archives). Järjestö pyrkii tukemaan maailmanlaajuisesti arkistojen hallinnoimien asiakirjojen säilymistä, hallintaa ja käyttöä:

”The International Council on Archives (ICA) is dedicated to the effective management of records and the preservation, care and use of the world's archival heritage through its representation of records and archive professionals across the globe”. (ICA, ei pvm)

ICA on luonut myös kansainväliset standardit mm. asiakirjojen ja toimijoiden kuvailuille. Asiakirjojen kuvailua varten on ISAD(G) (General International Standard Archival Description) (ICA, 2011) ja toimijoiden kuvailua varten ISAAR (CPF) (International Standard Archival Authority Record for Corporate Bodies, Persons and Families) (ICA, 2011). Molemmat standardit käsittelevät sitä, mitä asioita asiakirjojen tai toimijoiden kuvailussa tulisi käyttää tai tuottaa. Standardit ovat lähtöisin 1990-luvulta, jolloin internet ei vielä ollut kaikkien arkipäivää. Tämä myös on aiheuttanut muutospainetta standardeihin ja arkistokuvailua kohtaan ja tätä käsitellään lisää luvussa 3.2. Lisäksi ICA on luonut mm. toiminnan (ISDF -International Standard for Describing Functions) ja toimipisteiden (ISDIAH - International Standard for Describing Institutions with Archival Holdings) kuvailuun tarkoitettut standardit, mutta nämä eivät ole olleet käytössä Suomessa toistaiseksi kovinkaan laajasti, eikä näihin kiinnitetä tämän takia tarkempaa huomiota.

ISAD (G) ottaa kantaa paikkoihin vain epäsuorasti. Se ei määritä paikalle tai paikannimille omaa kuvailukohtaansa, vaan määrittää sen osaksi laajuutta/alaa ja sisältöä (scope

and content). Tämän lisäksi ISAD (G) ottaa kantaa paikkoihin vain kontekstin kautta. Kontekstin kautta luodaan yhteys toimijaan, jolla voi esimerkin mukaan olla syntymä- ja kuolinpaikka. Standardi kuitenkin kertoo, että tässä yhteydessä kannattaa myös hyödyntää ISAAR (CPF) –standardia, jossa nämä on määritelty tarkemmin.

ISAAR (CPF) käsittelee paikkoja suoremmin kuin ISAD (G). Standardissa toimijan kuvailulla on oma elementtinsä Paikka (Place). Tämän avulla on tarkoitus kuvata toimijan suhteet erilaisiin paikkoihin esimerkiksi sijainnin, elinpaikan, asumisen yms. osalta.

Molemmat standardit kuvaavat yleisellä tasolla, mitä asiakirjoista ja toimijoista tulisi kertoa. Standardit antavat myös esimerkkejä siitä, miten tämä tulisi tehdä. Esimerkki ISAAR (CPF) Place elementin käytöstä:

Birth: Krefeld, Germany (14 May 1899) Death: Edinburgh (17 March 1994)

United Kingdom, Royal Society (ICA, 2011)

Standardit eivät kuitenkaan ota kantaa siihen, miten tämä tulisi tietojärjestelmissä toteuttaa, mikä on todennäköisesti aiheuttanut ja tulee aiheuttamaan haasteita eri arkistoinstituutioissa, kun esimerkiksi metatietoja siirretään järjestelmistä toisiin, metatietojen syötössä on kirjoitusvirheitä tms. Tähän asiaan liittyen keskitytään tarkemmin aluvussa 2.3, jossa on tarkasteltu arkistolaitoksen nykytilannetta paikkoihin liittyen metatietojen osalta.

2.2. Arkistojen asiakkaiden tarpeet paikkoja koskevasta tiedosta

Arkistojen asiakkaiden palveleminen on keskittynyt pitkään henkilökohtaiseen ihmiseltä ihmiselle palvelemiseen. Palvelu on rakentunut usein asiakkaan ongelman ratkaisemisesta asiakkaan ja arkistoammattilaisen yhteistyönä. Tämä käy ilmi Duff et al. artikkelista, jossa mainitaan myös, että arkistojen käyttäjätutkimus on sangen rajoittunutta. Arkistot ovat tottuneet keräämään käyttäjätutkimusta ns. sisäisesti ilman, että tämä on ollut systemaattista. Tämä on oletettavasti myös osaltaan vaikeuttanut käyttäjätutkimusten tekemistä. (Duff, et al., 2006)

Arkistolaitos tekee nykyisin vuosittain asiakastyytyväisyyskyselyn, mutta tieto ei ole julkisesti julkaistua vaan se on käytettävissä vain arkistolaitoksen henkilökunnan toi-

mesta. Vuoden 2014 asiakastyytyväisyystutkimuksen tiivistelmässä kuitenkin nostetaan esiin seuraavia asioita:

Arkistolaitoksen tutkijasaleissa tapahtuvaan niin sanottuun fyysiseen asiakaspalveluun oltiin erittäin tyytyväisiä. Erityisesti asiakaspalvelun asiantuntevuus sai korkeat pisteet, yli 75 % oli siihen erittäin tyytyväisiä ja 17 % osittain tyytyväisiä. (Arkistolaitos, 2014)

Arkistolaitoksen sähköisiin palveluihin eivät asiakkaat olleet yhtä varauksettomasti tyytyväisiä kuin perinteisiin tutkijasalipalveluihin. (Arkistolaitos, 2014)

Näiden perusteella ei voida suoraan todeta, että paikkoihin liittyvät tarpeet olisivat asiakkaiden mielestä oleellisia. Voidaan kuitenkin todeta, että verkkopalveluiden käytettävyys ei tavalla tai toisella vastaa henkilökohtaista palvelua. Tähän vaikuttavat luonnollisesti monet asiat, mutta yleisesti voidaan todeta, että verkkopalveluiden tai sähköisten palveluiden kehittäminen paremmiksi olisi aiheellista. Pyrkimyksenä on kuitenkin tarjota tietopalveluita verkon kautta ja näiden tulisi olla keskeinen tapa toimia arkistolaitoksen strategian mukaan (Arkistolaitos, ei pvm).

Vuonna 2011 julkaistiin Suomessa SLS:n (Svenska Literatursällskapet i Finland) tilaama selvitys muistiorganisaatioiden asiakkaitten digitoitujen aineistojen tarpeista ja saatavuudesta (Hupaniittu, 2011). Selvitys keskittyy kirjastojen, arkistojen ja museoiden digitaalisiin aineistoihin ja näiden käyttäjäkuntaan. Selvityksessä on käsitelty 121 palveluiden käyttäjän vastausta. Arkistojen osalta tutkimuksessa esiin nousevia asioita voidaan soveltaa hyvin suureen osaan arkistoaineistosta, sillä arkistolaitoksen hallinnoimat digitaaliset aineistot ovat analogisesta aineistosta digitoituja eikä syntysähköisiä aineistoja ole arkistolaitoksen hallinnoimina toistaiseksi. Tämä on perusteltua, sillä paikkaan liittyvät tiedot ovat pääsääntöisesti yhtä päteviä aineistolle, oli kyseessä sitten paperinen käyttökappale taikka digitaalinen kuva.

Selvityksessä on tuodaan esille, että arkistolaitoksen Digitaaliarkisto (<http://digi.narc.fi/digi/>) on ollut kyselyyn vastanneiden palveluiden käyttäjien käytetyin verkkopalvelu. Arkistolaitoksen Digitaaliarkiston todetaan sen olevan käytettyydeltään hyvin monesta syystä heikko. (Hupaniittu, 2011) Heikkoutena mainitaan mm. hakemistojen ja indeksointien puutteellisuutta. Tällä tarkoitetaan kokonaisista arkistoyksiköistä

puuttuvia hakemistoja sekä yksittäisiin kuviin liittyvien tietojen puutteellisuutta. Arkistoyksikkö on alin arkistohierarkiassa esiintyvä taso, joka on käytössä nykyisin arkistolaitoksen Vakka-arkistotietokannassa. Tilanne voi kuitenkin olla se, että yksittäinen arkistoyksikkö pilkkoutuu useaan sataan kuvaan ja näillä kuvilla ei ole muuta metatietoa kuin suhde toisiinsa eli ns. järjestysnumero arkistoyksikön sisällä. Kuvassa 1 on esimerkki tästä.

DIGITAALIARKISTO

LINKIT

- Hakupalvelu
- Käyttöohjeet
- Viikon sisällä digitoidut esikirjat
- Digitointihanke 2009
- Digitointihanke 2010
- Digitoinnin eteneminen
- Kirjasto

Arkistoyksiköt:
Kirkonkirja 1890-1899 (1 Aa:9)

Digitoidut jaksot:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135
136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160	161	162	163	164	165
166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190	191	192	193	194	195
196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255
256	257	258	259	260	261	262	263	264	265	266	267	268	269	270
271	272	273	274	275	276	277	278	279	280	281	282	283	284	285
286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
301	302	303	304	305	306	307	308	309	310	311	312	313	314	315
316	317	318	319	320	321	322	323	324	325	326	327	328	329	330
331	332	333	334	335	336	337	338	339	340	341	342	343	344	345
346	347	348	349	350	351	352	353	354	355	356	357	358	359	360
361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
376	377	378	379	380	381	382	383	384	385	386	387	388	389	390
391	392	393	394	395	396	397	398	399	400	401	402	403	404	405
406	407	408	409	410	411	412	413	414	415	416	417	418	419	420
421	422	423	424	425	426	427	428	429	430	431	432	433	434	435
436	437	438	439	440	441	442	443	444	445	446	447	448	449	450
451	452	453	454	455	456	457	458	459	460	461	462	463	464	465
466	467	468	469	470	471	472	473	474	475	476	477	478	479	480
481	482	483	484	485	486	487	488	489	490	491	492	493	494	495
496	497	498	499	500	501	502	503	504	505	506	507	508	509	510
511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
526	527	528	529	530	531	532	533	534	535	536	537	538	539	540
541	542	543	544	545	546	547	548	549	550	551	552	553	554	555
556	557	558	559	560	561	562	563	564	565	566	567	568	569	570

Kuva 1. Digitaalinen arkiston arkistoyksikkö, jossa monta kuvaa.
(<http://digi.narc.fi/digi/slistaus.ka?ay=209404>)

Kuvassa 1 ”Digitoidut jaksot” tarkoittava arkistoyksikön (Kirkonkirja 1890-1899 (I Aa:9)) kuvia, joista jokainen numeroitu järjestysnumerolla (1-570). Tilanne voi olla kuitenkin parempi, mikäli arkistoyksikölle on tehty hakemisto. Tästä esimerkki kuvassa 2.



DIGITAALIARKISTO

LINKIT

- Hakupalvelu
- Käyttöohjeet
- Viikon sisällä digitoidut asiakirjat
- Digitointihanke 2009
- Digitointihanke 2010
- Digitoinnin eteneminen
- Kirjautu

Akaa

Vuodet: 1560 - 1579

Tunniste:

Arkisto: Asutuksen yleisluettelo

Sarja: Asutuksen yleisluettelot

Haanoja	1
Haihunkoski	3
Haudanniemi	5
Järvihaavisto	41
Kaulo	40
Konho	42
Kuhavuori	37
Kurisjärvi	28
Käyrälä	25
Lontila	26
Mustue	36
Nahkiala	16
Pätsiniemi	13
Riisikkala	27
Rättö	11
Sontula	38
Sotkia	33
Tarpia	10
Tiura	32
Toijala	20

Kuva 2. Digitaaliarkiston arkistoyksikkö, jossa hakemisto.
(http://digi.narc.fi/digi/hak_view.ka?hakid=44)

Kuvassa 2 arkistoyksikölle ”Akaa (1560-1579)” on luotu hakemisto, jossa ensimmäisessä sarakkeessa olevat tiedot alkavat samalla rivillä olevasta järjestysnumerosta (järjestysnumero toimii linkkinä kuvaan). Jo tästä esimerkistä voi hyvin havaita, että hakemisto on tuotettu paikannimien perusteella. Hakemistot ovat pääsääntöisestikin tuotettu

paikannimien perusteella, sillä hakemistoja tehdään vain niihin arkistoyksiköihin, joissa tälle on tarvetta. Erilaiset luettelomuotoiset arkistoyksiköt kuten kirkonkirjat, rippikirjat yms. usein tarvitsevat hakemistoa ja tällöin hakemisto luodaan juuri paikannimien perusteella. Arkistoyksiköt, joista ei muodostu useita kuvia, kuten esimerkiksi kartat harvoin tarvitsevat hakemistoja. Selvityksessä on mainittu, että karttamateriaali on hyvin käytettyä juuri digitaaliarkistossa (Hupaniittu, 2011). Arkistolaitoksen karttamateriaalia on kuvailtu, kuten muutakin arkistoaineistoa. Karttojen metatietoihin esimerkiksi ole liitetty muuta kuin tekstimuotoista tietoa. Esimerkki tästä kuvassa 3.



Kuva 3. Digitaaliarkistosta löydetty kartta.
(<http://digi.narc.fi/digi/slistaus.ka?ay=164822>)

Kuvassa 3 on Digitaaliarkistosta löydetty kartta ”3. Karta öfver Esbo socken. (Yleisk. Ia* 173:/- -)”, jonka nimi toimii linkkinä digitoituun kuvaan kyseisestä kartasta. Kartan kuvan voi siis löytää ilman, että tarvitsee selata useita satoja kuvia läpi. Kartalla ei kuitenkaan välttämättä ole muuta metatietoa kuin tuo nimeketeksti. Tämä saattaa aiheuttaa käyttäjälle ongelman, sillä karttaa ei esimerkiksi löydy hakusanalla ”Espoo”, sillä nimeketeksti on vain tekstimuotoisena eikä kone pysty identifioimaan sanaa ”Esbo” ruotsinkieliseksi vastineeksi sanalle ”Espoo”. Samaa termistön käyttöön liittyvä ongelma nostetaan myös Hupaniitun selvityksessä vapaasti vastattavissa kysymyksissä. Usea vastaaja on huomionnut ongelman, joka liittyy vain eksaktin termin käyttämiseen problematiikkaan. (Hupaniittu, 2011)

Hupaniitun selvityksessä selvitettiin loppukäyttäjien syitä siihen, miksi sähköistä aineistoa ei käytetty. Ongelmiksi koettiin mm. metatietojen puutteellisuus ja se, että käyttäjä ei löytänyt kiinnostavia aineistoja. Itse selvitys nostaa esille kaksi ongelmakohtaa, aineiston löytymättömyyden ja tietämättömyyden aineiston olemassaolosta. Lisäksi vas-

taajista 57/116 koki paikkaan liittyvän tiedon välttämättömäksi hakuehdoksi aineistolle. (Hupaniittu, 2011) Paikkatieto ei ole merkittävintä tietoa, jota ovat mm. aikatieto, tekijät, asiasanat ja nimekkeet, mutta se on kuitenkin tietoa, joita käyttäjät kokevat tärkeäksi. Vastauksissa on luonnollisesti otettava edelleen huomioon, että kyse ei ole vain arkistojen asiakkaista ja kysely on tehty koskien vain digitaalista aineistoa. Ei kuitenkaan voida sivuuttaa sitä tosiasiaa, että paikkatieto on relevanttia arkistoasiakkaille. Tätä väitettä tukee Hupaniitun selvitys ja arkistolaitoksen digitoitujen aineistojen luonne. Aineistojen luonteella tarkoitetaan esimerkiksi karttoja, joilla on vahva sidonnaisuus paikkoihin, mutta myös muita aineistoja, joissa paikkatiedolla on iso rooli aineiston löytyvyyden takia.

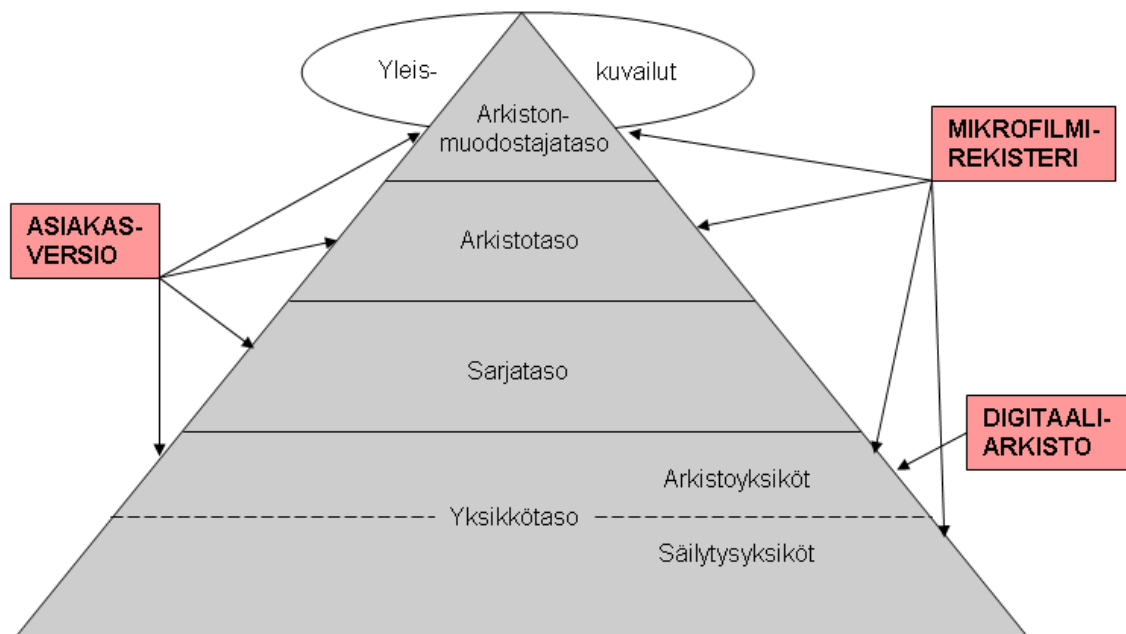
2.3. Arkistolaitoksen nykyjärjestelmät ja niiden sisältämä paikkatieto

Arkistolaitoksella on nykyisin kolme järjestelmää, jossa hallinnoidaan arkistoaineistojen metatietoja. Näistä laajin metatietojen määrällä mitattuna on Vakka-arkistotietokanta (<http://www.narc.fi:8080/VakkaWWW/EtuSivu.action>), joka käsittää arkistolaitoksen säilyttämien aineistojen kuvailu- ja luettelointitiedot pl. puolustushallinnon aineistot. Puolustushallinnon aineistojen metatietoja hallinnoidaan Aarre-arkistotietokannassa (<http://kronos.narc.fi/aarre/aarre.php>). Digitaaliarkistossa taas hallinnoidaan digitoitujen aineistojen metatietoja niiltä osin kuin ne eivät ole hallinnoitavissa Vakka-arkistotietokannan osalta. Näillä metatiedoilla tarkoitetaan edellisessä alaluvussa kuvattuja hakemistotietoja. Syy miksi tietojen hallinta on näin pirstaloitunutta, johtunee siitä, että Kansallisarkiston ja entisen Sota-arkiston yhdistyessä ei näiden tietojärjestelmiä yhdistetty, vaikka toimivat samassa loogisessa roolissa eli järjestelmänä, jolla hallinnoidaan arkistoaineistojen metatietoja. Digitaalisen aineiston hakemistotietojen toteuttaminen Vakka-arkistotietokantaan ei taas ole ollut mahdollista tai taloudellisesti järkevää, sillä kyseisen järjestelmä on peräisin 1990-luvulta ja tämän kehittämistä ei ole koettu sitä kautta perustelluksi.

Vakka-arkistotietokannan rakenne ja kentät noudattavat pitkälti Arkistojen kuvailu- ja luettelointiohjeita (Arkistolaitos, 1997). Kuvailu- ja luettelointiohjeet pohjautuvat taas ISAD (G) standardiin. ISAD (G) tai Arkistojen kuvailu- ja luettelointiohjeet eivät suoraan määritä järjestelmää tai tietorakennetta vaan antavat ennemminkin suuntaviivat tietorakenteelle.

Vakka-arkistotietokannan sisältämä kuvailu on hierarkista, kuten ISAD (G) ja Arkistojen kuvailu- ja luettelointiohjeissakin on määritelty. Hierarkian muodostavat arkistonmuodostaja, arkistot, sarjat ja arkistoyksiköt. Ylimpänä hierarkian tasona on arkistonmuodostaja. Arkistomuodostajalla voi yksi tai useampia arkistoja, joiden alla voi olla sarjoja. Sarjoja voi olla useamman tason verran. Alimpiin sarjoihin on liitetty arkistoyksiköitä, jotka ovat pienimpiä kuvailtuja kokonaisuuksia. Tämä rakenne on esitetty kuvassa 4.

VAKKA-arkistotietokannan rakenne ja kuvailutasot



Kuva 4. Arkistokuvailun hierarkia Vakka-arkistotietokannassa. (<http://www.narc.fi/tietokan.html>)

Arkistojen kuvailu- ja luettelointiohjeet määrittelevät kaikille tasoille yhteiset kuvailualueet ja elementit. Identifioinnin kuvailualueeseen kuuluva entiteetti 1.4 Maantieteellinen toiminta-alue on määritelty seuraavasti:

Maantieteellisellä toiminta-alueella tarkoitetaan sitä maantieteellistä tai hallinnollista aluetta, jonka arkistonmuodostajan toiminta kattaa tai on kattanut, tai arkistonmuodostajan kotipaikkakuntaa, sijaintipaikkaa tms. (Arkistolaitos, 1997)

Lisäksi määritelmä sisältää huomautuksen:

Maa tai paikkakunta, joka ei liity arkistonmuodostajan toimintaan, mutta josta aineistosta löytyy tietoa, esitetään hakusanana (kuvailuentiteetti 2.5) (Arkistolaitos, 1997)

Hakusanojen käyttö taas on määritetty seuraavasti:

Hakusanoja käytetään tehtäviä, erisnimiä (yhteisöjen, henkilöiden ja paikkojen), asiakirjatyyppien nimiä ja aineistojen tietosisältöä kuvaavia asiasanoja.

Hakusanat tulee esittää Yleisen suomalaisen asiasanaston standardoimalla tavalla. (Arkistolaitos, 1997)

Näiden pohjalta voimme päätellä, että paikannimistöä löytyy kahdesta eri kentästä ”Maantieteellisestä toiminta-alueesta” ja ”Hakusanoista”. Hakusanat voidaan rajata tämän työn osalta pois, sillä asiasanastoon pohjautuvat käsitteet linkitetään YSO:on (<http://finto.fi/ysa/fi/>) (Kansallisarkisto, 2013).

Tilanne Vakka-arkistotietokannan osalta ei kuitenkaan ole edellä esitetyn kaltainen, sillä paikannimistöä löytyy useammasta kentästä. Vakka-arkistotietokannassa arkistomuodostajilla on kaksi kenttää, johon paikannimistöä on ensisijaisesti tallennettu; ”PAIKKAKUNTA” ja ”MUU_ALUE” -kentät. Arkistoilla ja sarjoilla paikannimistölle ei ole osoitettu kenttää, mutta arkistoyksiköltä löytyy samat ”PAIKKAKUNTA” ja ”MUU_ALUE” -kentät. Kenttiin syötettävä tieto on vapaatekstiä ja sillä on ainoastaan pituusrajoitus.

Yhteenvedon voidaan todeta, että sekä ohjeistus, että toteutus ovat tietyllä tapaa aikaansa jäljessä. Vaikka ohjeistus on ylätasolla, eikä ota kantaa toteutustapaan, on sen pohjalta tehty toteutukset haastavia järjestelmä uudistusta ja tietojen siirtoa järjestelmästä toiseen suunniteltaessa. Kenttien vapaatekstimuotoisuus ja näiden täyttöohjeistus esimerkkeineen eivät juuri tue koneellisesti tehtävää tulkintaa ja se aiheuttaa yhden suurimmista haasteista tässä työssä. Tässä on esimerkki Arkistojen luettelointi- kuvailuohjeissa olevasta Maantieteellisen toiminta-alueen määrittämisestä:

Liperin kihlakunnan kruununvouti 1851-1944:

Kaavi 1851-1921, Kesälahti, Kitee, Kontiolahti, Kuusjärvi (nyk. Outokumpu), Lipperi, Outokumpu (ent. Kuusjärvi), Pielisensuu 1922-1944, Polvijärvi, Rääkkylä ja Säyneinen 1851-1921. (Arkistolaitos, 1997)

Oman haasteensa metatietojen läpikäymiseen tuo myös arkistolaitoksen hallinnollinen rakenne eli se, että se koostuu Kansallisarkistosta ja seitsemästä itsenäisestä maakunta-arkistosta. Vaikka nämä kaikki organisaatiot ovat käyttäneet samaa Vakka-arkistotietokantaa, on jokaisessa syötetty metatietoja tietokantaa itsenäisesti luettelointi- ja kuvailusääntöjä tulkiten.

2.4. Yhteenveto nykytilanteesta ja siinä havaituista ongelmista

Kuten edellisissä luvuissa tuotiin esille, on nykytilanteessa havaittavissa selkeitä ongelmia, jotka liittyvät arkistoaineistoihin ja niiden kuvailussa käytettyihin paikkatietoihin. Ongelmat juontuvat siitä olettamuksesta tai huomiosta, että asiakkailta on tarve hyödyntää paikkatietoa, kun he etsivät arkistoaineistoa. Nykytilanteessa arkistoaineistojen kuvailussa käytetty paikkatieto rajoittuu kuitenkin vain paikan nimiin tekstimuodossa ja paikannimistöä ei ole kyetty tuottamaan kaikilta osin tarpeeksi tarkalla tasolla. Tarkemmalle tasolle pääseminen voitaisiin korjata lisäämällä työvoimaa tuottamaan paikannimitietoa, mutta tämä ei ole tämän työn kannalta merkityksellistä. Vain tekstimuotoisen nimen korvaaminen ennalta määritetyllä termillä sen sijaan on asia, johon tämä työ ottaa kantaa ja kuvaa ratkaisumallin.

Ongelmat juontuvat osittain myös siitä, että paikannimistön tuottamisessa osaksi kuvailutietoa on käytetty osittain kansainvälisistä ja kansallisista ohjeista poikkeavia tapoja. Tämä taas johtuu osittain siitä, että nykyiset käytössä olevat järjestelmät ja niiden tietomallit eivät noudata yhteisiä ohjeistuksia. Eli jo järjestelmien rakennusvaiheessa on tehty ratkaisuja, jotka poikkeavat kansainvälisiin ja kansallisista ohjeista. Myös arkistolaitoksen fyysisesti eriytynyt organisaatio ja eri fyysisissä sijainneissa tehty työ ja sen laatu eroavat toisistaan.

3. Arkistolaitoksen järjestelmien ja palveluiden tavoitetila

Tavoitetilanteessa arkistolaitos on ottanut käyttöön arkistojen yhteisen AHAA-järjestelmän, jolla hallinnoidaan arkistoaineistojen kuvailu- ja luettelointitietojen metatietoja (Arkistolaitos, 2012). Paikkatietoon liittyvät metatiedot ovat osa kuvailutietoja. AHAA-järjestelmästä kuvailu- ja luettelointitiedon metatiedot välitetään arkistoaineistojen standardoiduissa siirtoformaateissa EAD ja EAC-CPF OAI-PMH rajapinnan avulla kirjastojen arkistojen ja museoiden yhteiseen palvelu ja tiedonhakuportaaliin Finnaan. (Kansallisarkisto, 2013) Finna rakentaa eri lähteistä saatavien metatietojen avulla oman sisäisen hakuindeksinsä, jota käytetään hakupalveluiden tuottamisessa (KDK, 2009).

Tässä luvussa tarkastellaan toimintaympäristöä, joka on vaikuttanut tavoitetilan syntymiseen sekä kuvattu tavoitetila niin palveluiden kuin toteutettavan AHAA-järjestelmänkin osalta.

3.1. Ohjaavat rakenteet ja sidosryhmät

Arkistolaitos on opetus- ja kulttuuriministeriön alainen viranomainen. Opetus- ja kulttuuriministeriö ohjaa arkistolaitoksen toimintaa mm. toimimalla sinä tahona, jonka kanssa arkistolaitos tekee tulossopimuksen. Opetus- ja kulttuuriministeriö ohjaa sen alaisten viranomaisten ja toimijoiden toimintaa rahoittamalla ja kehittämällä yhteisiä palveluita. Yksi merkittävimmistä arkistolaitosta ohjaavista tekijöistä on opetus- ja kulttuuriministeriö KDK-hanke (Kansallinen digitaalinen kirjasto -hanke). KDK-hankkeen tarkoituksena on lisätä kirjastojen, arkistojen ja museoiden digitaalisten aineistojen sekä kulttuuriperinnön merkitystä (Kansallinen digitaalinen kirjasto, 2015). Nykyisin KDK-hankkeesta voidaan ehkä ennemminkin puhua pysyvänä organisaationa, sillä sen pohjalta luotu toiminta on hyvin vakiintunutta ja sen toiminta on jatkunut yhtäjaksoisesti peräkkäisinä hankkeina. KDK:ta on esitelty enemmän seuraavissa alaluvuissa.

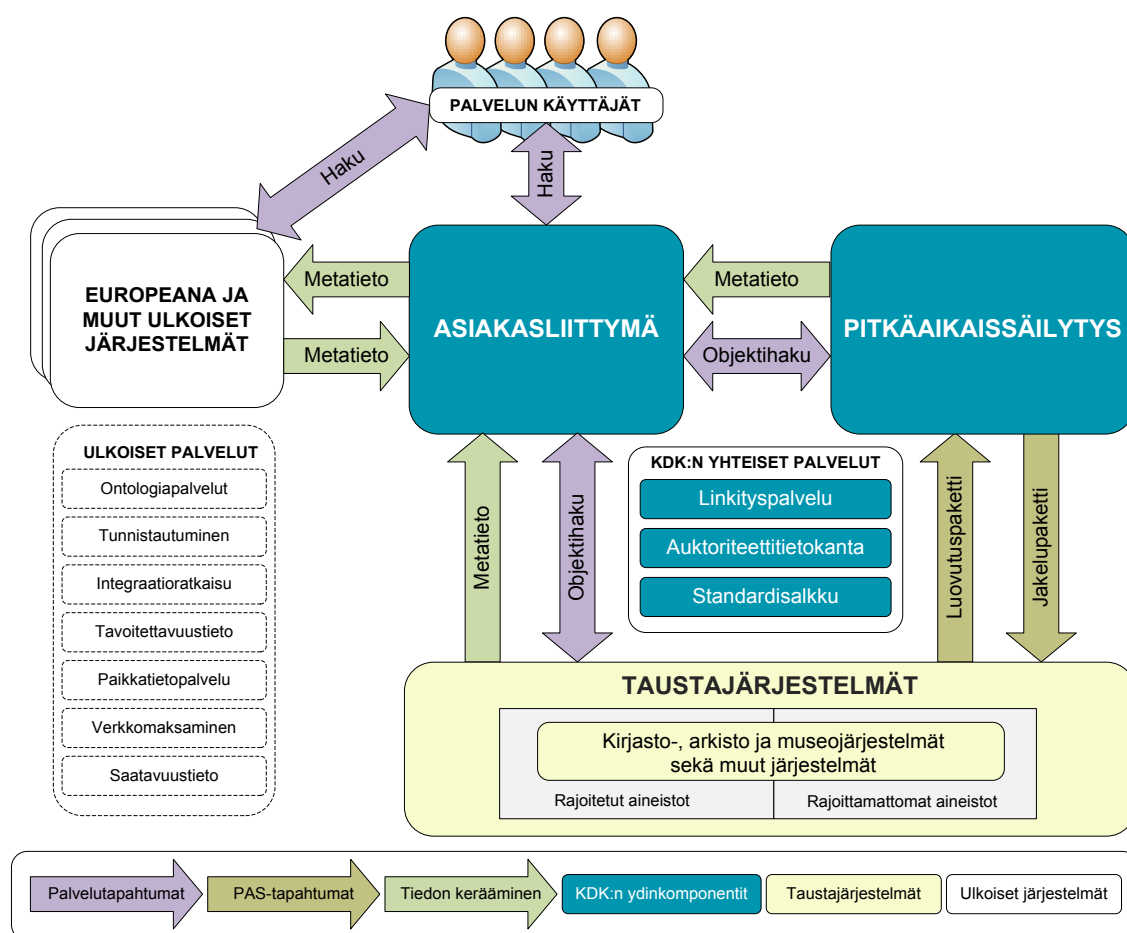
Muut tähän työhön merkittävästi vaikuttavat tahot ovat KDK:n alainen Finna (<https://www.finna.fi/>) sekä Kansalliskirjaston alainen Finto (<http://finto.fi/fi/>). Myös näitä molempia on esitelty tarkemmin seuraavissa alaluvuissa.

3.1.1. KDK – Kansallinen digitaalinen kirjasto

Kansallinen digitaalisen kirjaston toimintaa ohjaa opetus- ja kulttuuriministeriö. Kansallisen digitaalisen kirjaston toimintamallin ja sen toteuttamisen ohjauksen määrittää KDK:n kokonaisarkkitehtuuri. Kokonaisarkkitehtuuri määrittelee myös yhteisiä palveluita, tietosisältöjä, sovelluksia ja teknologiaa koskevat yhteentoimivuuden vaatimukset Kansalliseen digitaaliseen kirjaston osapuolille. (Kansallinen digitaalinen kirjasto, 2015)

Koska arkistolaitos on mukana KDK:ssa, asettaa KDK:n kokonaisarkkitehtuuri myös tiettyjä vaatimuksia ja ohjaavia sääntöjä sille. Toisaalta arkistolaitos pääsee myös suurimpana mukana olevana arkistona vaikuttamaan KDK:n ja sen kokonaisarkkitehtuuriin kehittymiseen jatkuvasti.

KDK kokonaisarkkitehtuurin kokonaiskuva on kuvattu seuraavasti:



Kuva 5. Kansallisen digitaalisen kirjaston arkkitehtuurin kokonaiskuva

Kuvassa 5 on esitetty KDK kokonaisarkkitehtuurin kokonaiskuva. Tummansinisellä on kuvattu KDK:n tuottamat yhteiset palvelut eli asiakasliittymä (Finna) ja pitkäaikaissäilytys (PAS). Taustajärjestelmillä tarkoitetaan kirjastojen, arkistojen ja museoiden omia järjestelmiä, joita ovat mm. arkistojen luettelointi- ja kuvailutietoja hallinnoivat järjestelmä ja näiden rajapinnat. Ulkoisina palveluina on mainittu mm. Ontologiapalvelut, joita nykyisin kansallisesti tuotetaan Finto-projektissa.

KDK kokonaisarkkitehtuuridokumentissa kuva 5 on avattu seuraavasti:

Kirjastojen, arkistojen ja museoiden kannalta keskeiset, priorisoidut aineistot digitoidaan ja saatetaan haettavaksi asiakasliittymän kautta. Osallistuva organisaatio hallinnoi omia aineistojaan taustajärjestelmissään, joista metatiedot haravoidaan asiakasliittymään. Asiakkaan tekemät haut kohdistetaan asiakasliittymään indeksoituun metatietokantaan ja tarvittaessa digitaalinen objekti haetaan taustajärjestelmästä asiakkaan käyttöön. Lisäksi KDK:n kautta metatietoja välitetään haettavaksi Euroopan digitaalisesta kirjastosta Europeanasta (Kansallinen digitaalinen kirjasto, 2012).

KDK kokonaisarkkitehtuuri määrittää siten, että osallistuvat organisaatiot hallinnoivat omia aineistojaan ja näiden metatietoja, mutta välittävät nämä myös yhteiseen asiakasliittymään (Finna), jossa on toteutettu yhteinen metatietohaku. Kokonaisarkkitehtuuri ei kuitenkaan ota kantaa tai mene niin syvälle toistaiseksi määrittämisissään, että se ohjeistaisi kaikkien metatietojen tuottamista tai hallinnointia. Se antaa vain rakenteet näiden välittämiseksi osapuolten välillä.

KDK kokonaisarkkitehtuuri määrittää siis ensisijaisesti asioita, jotka liittyvät sen itsensä tuottamiin palveluihin eli Finnaan ja PAS:iin. Näitä ovat mm. rajapintamääritykset sekä käytettävät metatietostandardit, formaatit yms., jotka liittyvät Finnaan tai PAS:iin. Tämä on erikseen määritetty KDK kokonaisarkkitehtuurin standardisalkussa (Kansallinen digitaalinen kirjasto, 2014). KDK kokonaisarkkitehtuuri määrittää myös ulkoiset palvelut ja näiden suhteen itseensä. Ontologiapalvelut on määritelty taulukon 1 mukaisesti.

Taulukko 1. KDK kokonaisarkkitehtuurin määritelmä ulkoisesta ontologiapalvelusta. (Kansallinen digitaalinen kirjasto, 2012)

Ontologioiden käyttöönoton ja ylläpidon palvelut	
Kuvaus	<p>Palvelu sisältää keskeisten kansallisten ontologioiden kehittämisen ja käytön palvelut. KDK-toimijoille tuotetaan kansallisen ontologiakirjaston ja -palvelimen (ONKI) tarjoamat palvelut sekä tarvittavat työkalut ontologiatekniikkaan:</p> <ul style="list-style-type: none"> ▪ Semanttisten portaalien toteuttamiseen tarkoitetut hakukoneet ja semanttiset linkittäjät. ▪ Metatiedon tuottamisen apuvälineet, kuten Saha-annotaatioeditori ja Poka-käsitetunnistin. ▪ Semanttisten portaalien keskinäisen tiedonvälityksen välineet. <p>Palvelun avulla voidaan siirtyä tiedon indeksoinnissa ja haussa nykyisestä avainsanatekniikasta semanttisesti rikkaampaan ontologiatekniikkaan, mikä mahdollistaa aiempaa käyttäjäystävällisemmän täsmätiedonhaun ja tietojärjestelmien semanttisen yhteismitallisuuden.</p>
Vastuu	Osallistuvat organisaatiot (ontologian ylläpito)
Sidosryhmät	ONKI-palvelusta vastaa vuoden 2012 syksyyn asti FinnONTO-hanke. Kansalliskirjasto neuvottelee VM:n kanssa palvelun pysyvän ylläpidon organisoinnista ja keskitetystä rahoituksesta.

KDK kokonaisarkkitehtuuri ottaa standardisalkun kautta kantaa siihen, miten ontologioita tulisi käyttää sisällönkuvailussa asiasanoituksen pohjalta rakennettujen ontologioiden osalta. Se ei kuitenkaan ota kantaa siihen, miten paikkatietoa tai paikkatieto-ontologioita tulisi soveltaa kuvailussa tai miten esimerkiksi ”käyttäjäystävällisempiä täsmätiedonhakuja” toteutettaisiin Finnassa. KDK kokonaisarkkitehtuuri on kuitenkin vaikuttanut esimerkiksi arkistojen yhteisen AHAA-järjestelmän suunnitteluun ja vaatimuksiin siten, että luetteloinnissa ja kuvailussa tullaan hyödyntämään entistä enemmän ontologisoitua tietoa. Ontologisoitua tietoa tullaan hyödyntämään ensisijaisesti asiasanoituksessa, mutta myös paikannimitiedon osalta SAPOn

(<http://dev.finto.fi/sapo/fi/>) avulla. SAPOn vahvuuksiksi arkistoinstituutioiden osalta on

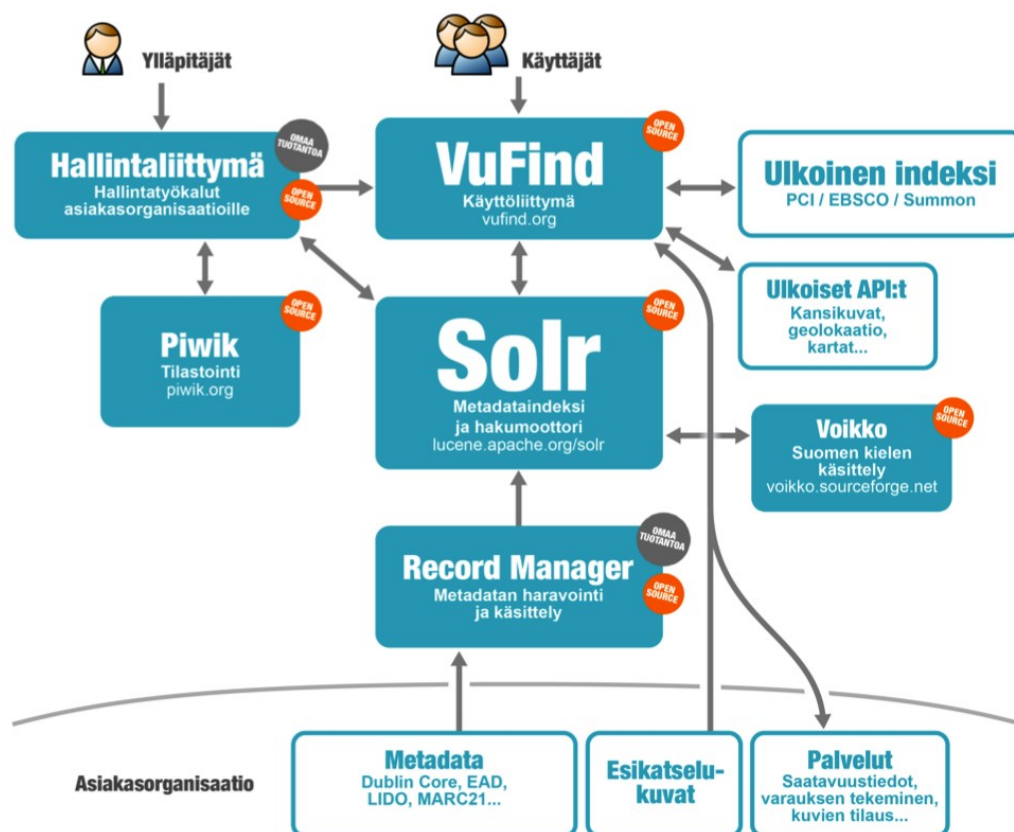
nähty erityisesti sen tarjoama ajallinen kattavuus, joka on jäänyt huomioimatta monissa muissa paikkatietopalveluissa. SAPOa on käsitelty lisää luvussa 4.3.

3.1.2. Finna

Finna on kirjastojen, arkistojen ja museoiden yhteinen verkkopalvelu, joka tarjoaa monipuoliset haut mukanaolevien organisaatioiden aineistojen hakuun. Lisäksi Finnan kautta on käytettävissä erilaisia aineiston käyttöön liittyviä palveluita kuten kuvien katselua, aineiston lainaamista yms. (Kansalliskirjasto, 2015)

Jotta yhteisen haun tarjoaminen eri kulttuurialojen sektoreiden (kirjastot, arkistot, museot) aineistoihin olisi mahdollista, on lähes välttämätöntä yhdenmukaistaa käytäntöjä. KDK kokonaisarkkitehtuurin standardisalkun puitteissa ohjeistetaan käyttämään tiettyjä teknologioita aineistojen metatietojen välittämiseksi yhteiseen hakupalveluun. Näiden metatietojen avulla Finna rakentaa oman sisäisen hakuindeksinsä, jossa eri sektoreiden tiedonsiirtoformaateista tiedot mapataan paikalliseen indeksiin. (Finna, 2015)

Paikallisen indeksin pohjalta Finna tarjoaa yhteisen haun kirjastojen, arkistojen ja museoiden aineistojen metatietoihin. Palvelu rakentuu useista eri avoimeen lähdekoodiin pohjautuvista ohjelmistoista, joista keskeisimpinä käyttöliittymän toteuttava VuFind ja hakupalvelimen toteuttava Solr ja metatiedon haravoinnin ja käsittelyn hoitava Record Manager. VuFind on kirjastojen kehittämä ja käyttämä haku ja palvelusovellus, Solr taas Apachen Lucene projektissa toteuttama hakupalvelin ja Record Manager kansalliskirjaston itse toteuttama sovellus. (Finna, 2015) Kuvassa 6 esitetty ohjelmistokokonaisuuden kuvaus. Arkistolaitos on kuvassa roolissa asiakasorganisaatio ja se tuottaa metatietoa Finnalle arkistojen metatietojen siirtoformaatilla EAD:lla.



Kuva 6. Finnan ohjelmistokokonaisuus. (Finna, 2015)

Toistaiseksi Finna ei ole ottanut huomioon juurikaan ulkoisia ontologioita tai niiden hyödyntämismahdollisuuksia hakupalveluissaan. Yhdeksi syyksi tähän on esitetty sitä, että varsin harva metatietoa tuottava organisaatio hyödyntäisi yhteisiä ontologioita kuvailussaan. Kyse on tässä yhteydessä ehkä enemmän ns. muna-kana -ilmiöstä. Kukaan ei hyödynnä ontologioita kuvailussa, koska kukaan ei hyödynnä niitä tarpeeksi konkreettisesti hauissa. Toisaalta kukaan ei voi hyödyntää tarpeeksi konkreettisesti hauissa kun kukaan ei hyödynnä ontologioita kuvailussa. Tästä johtuen arkistolaitos on pyrkinyt vaikuttamaan AHAA-hankkeen puitteissa Finnan suuntaan ontologioiden hyödyntämisen puolesta hauissa. AHAA:n valmistuessa tullaan jatkossa hyödyntämään erilaisia ontologioita kuvailussa ja toisaalta migraatiovaiheessa pyritään luomaan nykyisten metatietojen pohjalta linkityksiä olemassa oleviin ontologioihin. Vaikuttamisen tarkoituksena olisi saada otettua käyttöön kohtuullisen yhtäaikaaisesti metatiedon tuottaminen ontologisoidun tiedon osalta, kuin myös hakujen toteuttaminen ontologisoidun tiedon pohjalta.

3.1.3. Finto

Finto on palvelu sanastojen, ontologioiden ja luokitusten julkaisua ja käyttöä varten. Finto palvelua on kehitetty Finto-projektissa, joka on Kansalliskirjaston, Opetus- ja kulttuuriministeriön sekä valtionvarainministeriön yhteisprojekti. Finto palvelun ja projektin juuret juontavat Aalto yliopiston ja Helsingin yliopiston yhteiseen FinnONTO-hankeeseen, joka oli käynnissä vuosina 2003-2012. (Finto, 2015) FinnONTO-hankkeen visiona oli kehittää avoin perusta) suomalaisen semanttisen webin infrastruktuurille (Semantic Computing Research Group (SeCo), 2014). Karkeasti voidaan arvioida hankkeen onnistuneen luomaan pohjan suomalaiselle semanttisen webin infrastruktuurille, sillä tämän pohjalta syntyi Finto-projekti ja sitä myötä Finto-palvelu.

Itse palvelu tarjoaa hyödyntäjilleen erilaisia ontologioita ja näiden selailuun käyttöliittymän sekä avoimet rajapinnat. (Finto, 2015) Lähtökohtaisesti uudistuvassa arkistokuvailussa ja sitä toteuttavassa AHAA-järjestelmässä tullaan hyödyntämään näitä rajapintoja tarjoamalla kuvailijoilla mahdollisuus valita kohteita ontologioista. Asiasanoituksen osalta tämä tapahtuu esimerkiksi YSO:sta ja aikaan sidottujen paikkojen osalta SAPOsta (Kansallisarkisto, 2013). Toistaiseksi YSO on ylläpidetty ja tuettu ontologia, kun taas SAPO julkaistuna vasta kehitysvaiheessa.

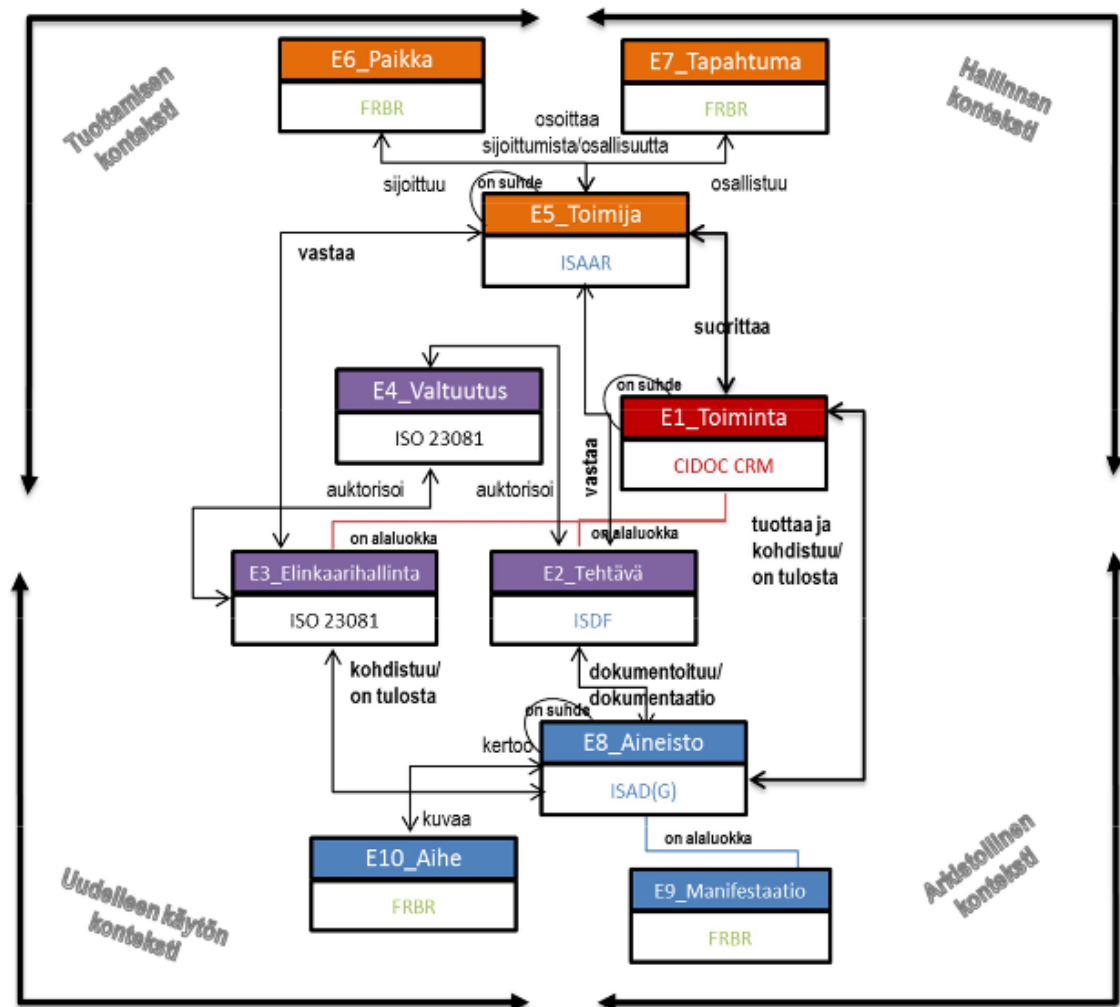
3.2. AHAA tietorakenne

AHAA-järjestelmän tietorakenne tulee perustumaan arkistokuvailun kansalliseen käsitemalliin. Arkistokuvailun kansallisen käsitemallin kehittäminen aloitettiin Kansallisarkiston toimesta vuonna 2012 ja se tavoitteena oli uudistaa arkistokuvailua muuttuneen toimintaympäristön johdosta. Tavoitteena käsitemallilla on toimia perustana elinkaaren eri vaiheissa olevan analogisen, digitoidun tai alkujaan digitaalisen aineiston arkistokuvailun metatietomalleille. Työ on jatkunut aina nykypäivään asti, mutta työn tavoitteessa on tehty se muutos, että kansallisia kuvailusääntöjä ei uudisteta ennen kuin kansainväliset kuvailusäännöt on päivitetty. Kansainvälinen työ jatkuu vuoden 2016 ICA (The International Council on Archives) yleiskokoukseen, jossa tulokset julkaistaan. (Kilkki & Henttonen, 2015)

Nyt julkaistu käsitemalli on vielä luonnosvaiheessa, mutta se tulee hyvin todennäköisesti olemaan yhdenmukainen uudistuvan kansainvälisen käsitemallin kanssa. Tämä oletamus voidaan tehdä, sillä Suomesta on edustus kansainvälisen käsitemallin uudistamis-

työssä. Yksi vaihtoehto on myös se, että kansallista mallia ei tarvita, mikäli kansainvälinen malli on kansallisetkin tarpeet tyydyttävä. Käsittemallin luonnoksen tarpeellisuutta voidaan perustella viestinnän välillä eri kulttuurisektoreiden (kirjastot, arkistot, museot) välillä. Sitä hyödynnetään myös KDK:n piirissä tehtävän eri sektoreiden kuvailun harmonisointiin liittyvässä työssä. (Kilkki & Henttonen, 2015)

Kuvassa 7 on esitetty arkistokuvailun kansallinen käsitemalli. Kuvan vasemmassa yläkulmassa on kuvattu entiteetti E6_Paikka ja tämän suhteet käsittemallin muihin entiteetteihin.



Kuva 7. Arkistokuvailun kansallinen käsitemalli ja elinkaaren jatkumon kontekstit. (Kilkki & Henttonen, 2015)

Entiteettiä E6_Paikka kuvataan käsitemallissa taulukon 2 mukaisesti. Taulukossa 2 on kuvattu kohteen eli ”E6_Paikan” määritelmä, kuvaus, taustastandardi, esimerkit ja suhteet käsitemallissa.

Taulukko 2. Entiteetin Paikka_E6 määritelmätaulukko (Kilkki & Henttonen, 2015)

Määritelmä	Representoi fyysistä tai virtuaalista E5_Toimijan toimintakontekstiin liittyvää paikkaa.
Kuvaus	<p>Paikka ymmärretään laajasti historialliseksi tai nykyiseksi maantieteelliseksi tai hallinnolliseksi alueeksi, fyysiseksi tai virtuaaliseksi tilaksi tai paikaksi, jolla on maantieteellisen koordinaatiston avulla ilmastavissa oleva sijainti (sijaintipaikka).</p> <p>Paikka on osa sitä sosio-kulttuurisen kontekstin kuvailua, jossa toimija on toiminut ja josta aineisto näin sisältää tietoa.</p> <p>Paikka on modulaarisessa kuvailumallissa lähtökohtaisesti erillinen kuvailuentiteetti ja implementoinnissa auktoriteettitietoa, joka kuvataan ontologiana.</p> <p>Paikalla voi olla suhde myös toimintaan ja tapahtumaan, mutta käsitelmallisä tämä suhde rakentuu implisiittisesti toimijan kautta.</p>
Taustastandardi	<p><i>Functional Requirements for Bibliographic Records</i> (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009)</p> <p>The entity defined as place encompasses a comprehensive range of locations: terrestrial and extra-terrestrial; historical and contemporary; geographic features and geo-political jurisdictions.</p>
Esimerkit	Paikka voi olla kunta, jonka alueella toimija asuu tai karttakoordinaatein ilmaistava toimijan pääkonttorin sijaintipaikka; hallintoalue, jonka toimijan toiminta eri aikoina on kattanut; tai virtuaalinen tila, jossa toimija on osallistunut kokoukseen tai keskusteluun.
Suhteet	RS_6a = E6 ja E5_Toimija välillä: osoittaa sijoittumista. Osoittaa toimijan historiallisen toimintakontekstin kiinnittymistä paikkaan.

Taulukossa kohdassa ”Kuvaus” on määritetty kohta, jossa todetaan, että paikka on lähtökohtaisesti erillinen kuvailuentiteetti ja sen ajatellaan olevan auktoriteettitietoa, joka on kuvattu ontologisesti. Tämä toimii sinä kantavana ajatuksena, joka on eniten vaikuttanut AHAA-järjestelmän tietomalliin paikkatiedon osalta. Tämä on kuvattu tarkemmin luvussa 3.3.

AHAA-järjestelmä toteuttaa rooliaan arkistokuvailun metatietojen hallinnointityökaluna ja siinä käytetty metatietomalli pohjautuu arkistokuvailun kansalliseen käsitemalliin. AHAA-järjestelmän metatietomalli ei ole täysin yhdenmukainen käsitemallin kanssa, sillä se on jouduttu määrittelemään ennen kansallista käsitemallia AHAA-hankkeen aikataulun ja kansallisen kuvailutyöryhmän työskentelyn ollessa päällekkäiset, eivätkä peräkkäiset.

3.3. Paikkaan liittyvien metatietojen hallinta AHAA-järjestelmässä

AHAA-järjestelmässä paikkatiedon hyödyntämisessä käytetään Finto-palvelun tarjoamaa ONKI-selectoria, jonka avulla voidaan valita paikkaontologioiden kohteita ja luoda linkityksiä näihin. AHAA-järjestelmään ei lähtökohtaisesti kopioida ontologioita vaan ainoastaan niistä löytyviä URI:n (Uniform Resource Identifier) avulla identifioituja kohteita. Kohteista AHAA:n puolella kopioidaan vain välttämättömät tiedot, joiksi on ajateltu URI ja kohteen label eli toisin sanoen nimeke. Nimekettä tarvitaan AHAA-järjestelmässä vain arkistohenkilökunnan suorittamiin hakuihin, joiden ei tarvitse olla niin kehittyneitä kuin asiakkaille tarjottavien hakujen. Aineiston metatietoja välitettäessä AHAA-järjestelmästä ulos, voidaan paikkatiedon metatietoja täydentää ns. lennosta tai vaihtoehtoisesti jättää täydentämättä. Finnaan metatietoja välitettäessä paikkatietoon liittyviä tietoja ei lähtökohtaisesti täydennetä vaan välitetään vain kohteiden URI:t ja labelit. (Kansallisarkisto, 2013)

Paikkatiedon hallinnan osalta AHAA:ssa ei ole vielä tehty lopullista toteutusratkaisua, sillä nykyisin (tilanne 31.5.2015) esimerkiksi SAPO-ontologiasta ei ole julkaistu tuotantoversiota. Lisäksi tiedot, jotka mielletään paikkatiedoksi, mutta joita ei voida rinnastaa SAPOsta löydettyihin kohteisiin on hallinnoitava jollakin muulla tapaa kuin edellä on esitetty. Nykyisten vaatimusten mukaan myös AHAA-järjestelmän sisäinen paikkatietojen hallinta tulee olla mahdollista AHAA-järjestelmässä.

Tavoitteena on, että Finna täydentää sisäistä hakuindeksiään tarvittavin määrin URI:ien takaa löytyvillä metatiedoilla esimerkiksi paikkatiedon osalta koordinaateilla tai vieras-kielisillä nimekkeillä. Tämä on perusteltua, sillä Finnaan metatietoa välitetään kymmenien tai satojen järjestelmien kautta ja on kokonaisvaltaisesti järkevämpää tehdä tarvittavat täydennykset keskitetysti yhdessä paikassa kuin jokaisessa järjestelmässä erikseen. Vaihtoehtoisesti Finna voisi tarjota ontologiapohjaisia selailumahdollisuuksia paikkatieto-ontologian pohjalta. Ontologioiden hyödyntämismahdollisuudet ovat laajoja ja näitä on käsitelty enemmän luvussa 4.2.

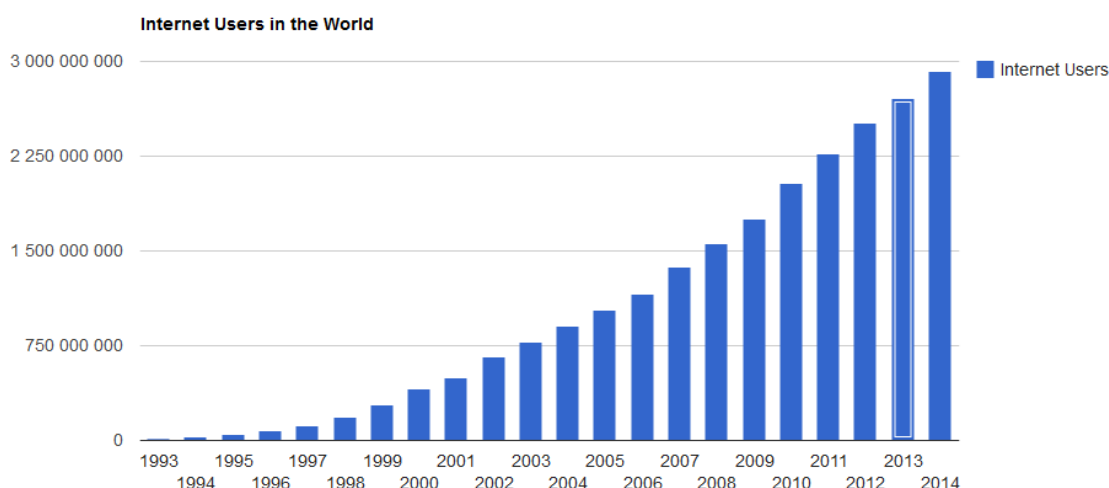
4. Teknologia ja sen hyödynnettävyys

Tässä luvussa tarkastellaan työssä kuvatus ratkaisun mahdollistavaa teknologiaa sekä sen hyödynnettävyyteen liittyvää taustatietoa. Luku on siten teoriapainotteinen. Aluksi esitellään läpi semanttiseen webiin ja avoimeen linkitettyyn dataan liittyviä teknologioita. Luvun jälkimmäinen puolisko esittelee teknologian pohjalta tuotettuja sovellutuksia sekä teknologian tarjoamia mahdollisuuksia. Luvuntarkoituksena on avata kuvattuun ratkaisuun ja asetettuun tavoitettiin liittyviä teknologioita ja niiden mahdollistamia hyötyjä.

4.1. Teknologiat ratkaisun taustalla

4.1.1. WWW – World Wide Web

WWW sellaisena kuin me se miellämme on muuttanut maailmaa hyvin paljon viimeisten vuosikymmenten ajan. WWW on muuttanut kommunikointitavat, joita ihmiset käyttävät sekä sen miten liiketoimintaa harjoitetaan (Antoniou & van Harmelen, 2004). Tämä on ollut tilanne jo vuosia, eikä WWW:n käytön kasvuvauhti ole hidastunut juuri lainkaan. Internet ja WWW ovat arkipäivää lähes kaikessa tekemässämme. Kuvassa alla on esitetty internetin käyttäjämäärän kasvu 1990-luvun alkupuolelta nykypäivään saakka. (Internet Live Stats, 2015)



Kuva 8. Maailman internetin käyttäjien määrän kehitys 1993-2014. (Internet Live Stats, 2015)

Nykyisessä WWW:ssä voidaan nähdä kuitenkin ongelma. Googlen kautta voidaan hakea, mutta haku kohdistuu yksittäisiin sivuihin; HTML dokumentteihin. Tämän kaltai-

sessä hakutavassa on kuitenkin puutteensa. Saadaan paljon tuloksia, mutta vain osa on relevantteja, osumia ei saada lainkaan, hakutulokset ovat hyvin riippuvaisia käytetystä termistöstä ja/tai hakutulokset ovat yksittäisiä sivuja. (Antoniou & van Harmelen, 2004) Kone ei ymmärrä hakemaansa, eikä välttämättä osaa esittää hakijalleen relevantteja hakutuloksia. Esimerkiksi hakusanalla ”golf” voidaan saada hakutuloksia liittyen urheilulajiin golf tai Volkswagen yhtiön automallin Golf. Vastaavantyypisesti käyttämällä hakusanaa ”hilla” ei välttämättä saada hakutuloksia kohteista, joissa on käytetty termiä ”lakka” tai ”suomurain”, vaikka kyse on lopulta samasta tai vähintäänkin relevantista termistä. Yhtenä ratkaisuna tämän kaltaisen ongelman ratkaisuun WWW:ssä on esitetty semanttista webiä, jossa pyrkimyksenä on mahdollistaa koneiden tekemään päättelytyötä.

4.1.2. Semanttinen web

Semanttisella webillä on Tim Berners-Leen et al. mukaan tarkoitettu internetiä, jossa tieto ei ole vain ihmisten ymmärrettävässä muodossa, vaan myös koneiden ymmärtämässä muodossa (Berners-Lee, et al., 2001). Jo tuolloin, noin 15 vuotta sitten semanttisen webin ennustettiin olevan alusta, jossa tietokoneet voivat itsenäisesti tehdä monimutkaisempia ihmisen antamia tehtäviä kuin vain seurata linkkejä sivulta toiselle.

W3C (World Wide Web Consortium) määrittelee semanttisen webin tärkeimmiksi tavoitteiksi sen, että koneet voisivat tehdä hyödyllisempiä töitä ja voitaisiin kehittää järjestelmiä, jotka voisivat olla luottamuksellisessa vuorovaikutuksessa keskenään. Edistääkseen tavoitteita on määritetty erilaisia teknologioita tukemaan tätä. Näitä ovat mm. RDF (Resource Description Framework), SPARQL (Simple Protocol and RDF Query Language), OWL (Web Ontology Language) ja SKOS (Simple Knowledge Organization System). (W3C, 2015) Näistä keskeisin ymmärtää työn kannalta on RDF.

RDF – Resource Description Framework

RDF on malli, jonka avulla voidaan määritellä ja kuvata resursseja WWW:ssä. Sen tarkoituksena on kuvata asioita ennalta määritellyllä tavalla, jotta koneet voisivat tulkita tietoa helpommin. RDF:n perustana toimii kohteiden tunnistaminen URI:ien (Uniform Resource Identifier) avulla. RDF:n avulla voidaan ilmaista asioita ns. kolmikoiden avulla. Kolmikko koostuu subjektista, objektista ja predikaatista. (W3C, 2015) Yksikertainen esimerkki voisi olla seuraavan kaltainen:

”Matti” (subjekti)

”asuu osoitteessa” (predikaatti)

”Kalliomäki 5” (Objekti)

Kolmikon kohteet voidaan identifioida URI:ien avulla, mutta vain predikaatin identifioiminen URI:lla on pakollista. RDF määrittää tietyn joukon ominaisuuksia, joita sen avulla voidaan kuvata. (W3C, 2015)

URI

URI on tunniste, jolla voidaan identifioida kohde tai resurssi WWW:ssä. URI on yläkäsite URLille (uniform resource locator) ja URN:lle (uniform resource name) ja sitä voidaan käsitellä joko URL:na, URN:na tai molempina. (Berners-Lee, 1998)

SPARQL

SPARQL on kyselykieli, jolla voidaan tehdä kyselyitä RDF -muotoiselle datalle. Kyselyitä voidaan tehdä riippumatta siitä, onko data alkuperäisesti RDF muodossa vai julkaistu RDF muodossa esimerkiksi väliohjelmistossa. (W3C, 2015)

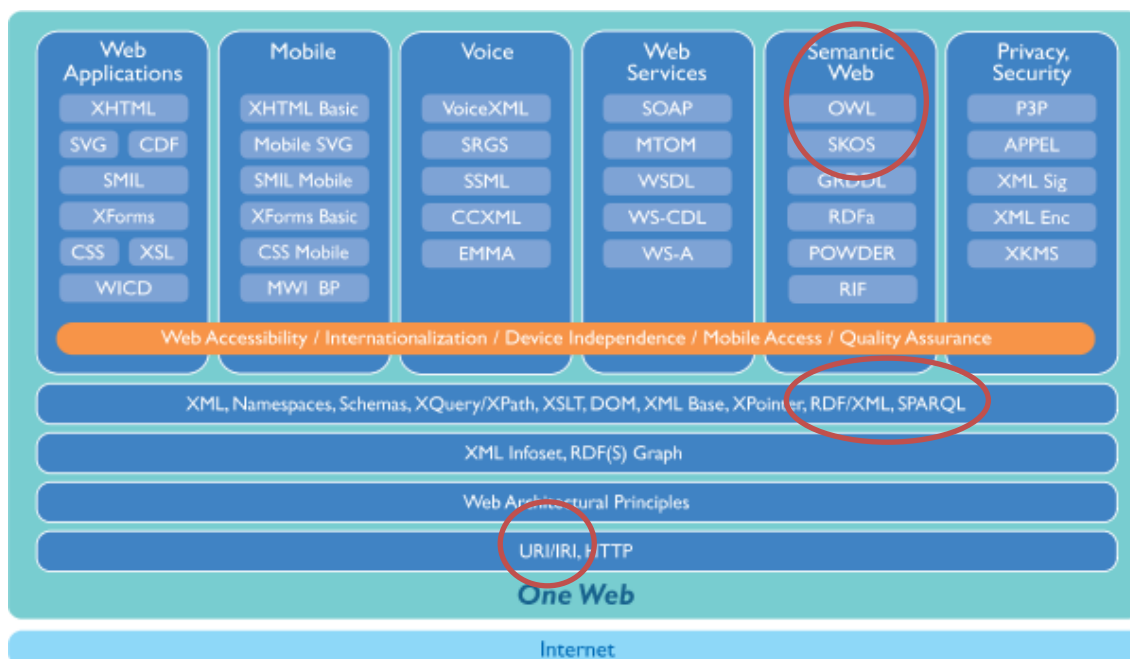
OWL

OWL on nimensä mukaisesti ontologioiden kuvauskieli. Sen tarkoituksena on mahdollistaa kohteiden kuvaaminen koneymmärrettävästi. Se on monimuotoisempi kuin RDF ja se mahdollistaa sitä myötä rikkaammat ontologioiden kuvausmahdollisuudet. (W3C, 2015)

SKOS

SKOS on datamalli yksinkertaisille datakokoelmille. SKOS:n avulla voidaan ilmaista esimerkiksi erilaisia thesauruksia, sanastoja, taksonomioita tai luokituksia.

Kuvassa 9 on esitetty näiden eri teknologioiden suhteet toisiinsa W3C teknologiapiinossa.



Kuva 9. W3C Technology Stack illustration (<http://www.w3.org/2004/10/RecsFigure-Smaller.png>)

Kuvassa 9 on ympyröity edellä esiteltyt teknologiat. Voidaan sanoa, että RDF ja SPARQL toimivat perustana semanttiselle webille ja OWL tuo ontologioiden mallinnukseen rikkaamman työkalun kuin RDF, kun taas SKOS on datamalli, jolla voidaan mallintaa yksinkertaisia ontologioita. URI:t taas toimivat yksilöinnin apuvälineenä kaikilla tasoilla.

4.1.3. Ontologiat

Ontologia termiä on käytetty filosofiassa kuvaamaan mitä on olemassa ja miten olemassa olevaan voidaan kuvata. Tätä määritelmää on käyttänyt myös Gruber oman määritelmänsä pohjana, jonka mukaan ontologia on eksplisiittinen määritelmä käsitteellistämiseksi. Ontologialle on ominaista kuvata siinä esiintyvien entiteettien luokkia, suhteita, toimintoja ja kohteita. (Gruber, 1993)

Tietojenkäsittelyssä ontologia termin käyttäminen on vakiintunut viime vuosien aikana. Ontologialla tarkoitetaan mallinnusta, jossa sen sisältäminen entiteettien väliset suhteet on kuvattu määrämuotoisesti siten, että näiden avulla voidaan tehdä konepäättelyä. (Wikipedia, 2015) Semanttisen webin teknologiat ovat siis luotu, jotta tämä olisi mahdollista.

Käsitys ontologioista on muovautunut arkistojen näkökulmasta enemmän määrämuotoisiin luokitteluihin ja yhteiskäyttöisiin tietovarantoihin, joiden kohteet ovat yhteisesti määriteltyjä ja sovittuja. Tätä käsitystä on varmasti vahvistanut myös Finton rooli ja Finton ylläpitämien ontologioiden rooli. Voidaan myös ajatella, että ontologia termiin liittyy vahvasti myös käsitys avoimuudesta ja yhteiskäyttöisyydestä.

4.1.4. Avoin linkitetty data

Avoimella linkitetyllä datalla on monta määritystä ja määritys vaihtelee hieman lähteestä riippuen. Yksi käytetyimmistä määritelmistä on kuitenkin Tim Berners-Leen määritelmä (2006). Hän määrittelee linkitetyn datan olevan dataa, jolla voit löytää lisää dataa. Linkitetyn datan neljäksi perustaksi hän määrittelee seuraavat kohdat:

1. *Käytä URI:ja nimeämiseen*
2. *Käytä HTTP URI:ja, jotta niitä voidaan ”seurata” tai ”katsoa mitä sen takaa löytyy”*
3. *Mikäli URI:a ”seuraa” tarjoa sen takaa hyödyllistä tietoa standardeja noudattaen (RDF*, SPARQL)*
4. *Sisällytä linkejä muihin URI:iehin, jotta etsijä voi löytää enemmän (Berners-Lee, 2006)*

Mikä sitten tekee linkitetystä datasta avointa linkitettyä dataa? Tim Berners-Lee on jatkanut linkitetyn datan määritelmää avoimeksi linkitettyksi dataksi sillä perusteella, että data on avointa kun se on julkaista avoimella lisenssillä. Avoimuudella tarkoitetaan tässä yhteydessä sitä, että sen käyttö ei ole maksullista. (Berners-Lee, 2006)

Avoin linkitetty data voidaan Tim Berners-Leen mukaan jaotella viiteen eri tähtiluokkaan. Linked Data Finland on jatkanut tätä luokitusta kahdella lisätähdellä seitsemään tähteen. Luokitukset on kuvattu seuraavasti:

★ *Available on the web (whatever format) but with an open licence, to be Open Data*

★★ *Available as machine-readable structured data (e.g.*

excel instead of image scan of a table)

★★★

as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★

All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★

All the above, plus: Link your data to other people's data to provide context (Berners-Lee, 2006)

ja Linked data Finlandin laajennus:

★★★★★★

Provide your data with a schema and documentation so that people can understand and re-use your data easily.

★★★★★★★

Validate your data and denote its provenance so that people can trust the quality of your data. (Linked Data Finland, ei pvm)

Yhteenvetona voidaan todeta, että avoimen linkitetyn datan edellytyksenä ovat monet semanttisen web teknologiat. Niiden hyödyntäminen yksilöivien tunnisteiden (URI) kanssa ja julkaisemalla tietoa avoimesti saavutetaan avointa linkitettyä dataa. Arkistojen ja monien muiden asioita tai kohteita kuvailevien organisaatioiden (esimerkiksi museot ja kirjastot) osalta voidaan päätellä, että avoin linkitetty data tulee olemaan jatkossa entistä suuremmassa roolissa. Väittäminen perustuu esimerkiksi siihen, että kaikilla näillä tahoilla on tarve tunnistaa henkilöitä, organisaatioita, tapahtumia, paikkoja yms. joiden kuvaaminen ja määrittäminen yksin ei ole tehokasta tai järkevää. On mielekkäämpää esimerkiksi viitata jo määritettyyn Talvisotaan kuin määrittää se itse. Samansuuntaisesti asian on tuonut ilmi Summers & Salo artikkelissa, jossa todetaan, että kulttuuriperintösektorilla (kirjastot, arkistot, museot) on viime aikoina ollut kiinnostus kohti semanttisen webin mahdollistamia asioita (Summers & Salo, 2013).

4.2. Ontologioiden hyödyt

Ontologioiden hyödyt eivät yksittäin esitettyinä ole selkeästi parempia, kuin muiden teknologioiden. Ontologioiden ominaisuudet yhdessä kuitenkin tekevät niistä erityisen kokonaisuuden. Ontologioiden keskeisimmät ominaisuudet, jotka tekevät niistä hyödyllisiä ovat käsitteellinen mallintaminen, joustava skeemakehitys, suora vuorovaikutus, uudelleenkäytettävyys, parhaat käytännöt, webin sääntöjen noudattaminen, formaalisuus ja päättely. (Janowicz & Oberle, 2009)

Käsitteellisen mallintamisen hyötynä on se, että ontologiat voidaan mallintaa helposti ymmärrettävien luokkien, ominaisuuksien, sääntöjen ja aksioomien sekä instanssien avulla. Joustavan skeemakehityksen avulla luokkien, ominaisuuksien, sääntöjen ja aksioomien sekä instanssien muokkaaminen on mahdollista myös ns. ajon aikana, eikä vain esimerkiksi suunnitteluvaiheessa. Suora vuorovaikutus mahdollistaa em. ominaisuuksien avulla helpomman ontologioiden muuttamisen ja jatkohyödyntämisen esimerkiksi visualisoinnilla. Uudelleenkäytettävyys on helpompaa, kun ontologia on käsitteellisesti mallinnettu. Parhaat käytännöt ovat muotoutuneet ontologiakehityksessä vuosien varrella ja auttavat uusien ontologioiden kehittämisessä. Webin sääntöjen noudattamisella tarkoitetaan esimerkiksi RDF:n ja SPARQL:n hyödyntämistä, jotka sopivat hyvin yhteen webin kanssa, kuten W3C teknologiapinossakin on kuvattu (Kuva 9. W3C Technology Stack illustration). Formaalisuudella tarkoitetaan ontologioiden logiikkaa, joka mahdollistaa päättelyn, jonka avulla voidaan tarkistaa mallinnettujen asioiden oikeellisuutta tai luoda ”uutta tietoa”. (Janowicz & Oberle, 2009)

Se, onko ontologioiden hyödyntäminen kannattavaa esimerkiksi taloudellisesti, on riippuvaista monestakin asiasta. Tähän vaikuttavat esimerkiksi teknisempien henkilöiden osaamisen ja kiinnostuksen kehittyminen aihealueeseen. (Janowicz & Oberle, 2009)

Suomessa semanttiseen webiin liittyvää koulutusta on tarjottu TKK:lla (Teknillinen korkeakoulu) ja Aalto yliopistossa jo useita vuosia. Koulutuksen kanssa tiivistä yhteistyötä on tehnyt semanttisen laskennan ryhmä, joka koostuu Aalto yliopiston ja Helsingin yliopiston väestä. Voidaan päätellä, että vuosien varrella kertyneen osaamisen johdosta, on perustettu mm. Finto-projekti, jonka henkilöstö suurelta osin on lähtöisin semanttisen laskennan ryhmästä ja/tai Aalto yliopiston koulutusohjelmista.

Ontologioiden hyödyntämisellä voidaan ainakin siis teoriassa saavuttaa hyötyjä. Tämän työn ratkaisun avulla voidaan mahdollistaa ontologisoituun tietoon perustuva tiedonha-

keminen. Vuonna 2007 tehdyn arvioinnin perusteella on tultu siihen tulokseen, että suurin osa ontologioihin liittyvistä toimenpiteistä on keskittynyt ontologioiden suunnitteluun. (Ahmad & Colomb, 2007) Itse ontologioiden hyödyntämiseen esimerkiksi hakujen tukena ei ole niin laajamittaista tutkimusta saatavilla. Tähän kuitenkin perehdytään seuraavissa alaluvuissa.

4.2.1. Ontologioiden hyödynnettävyys tiedon hakemisessa

Kuten hakemisessa usein, on myös ontologioihin tai semanttisen webin teknologioihin perustuvan haun oltava käyttäjälleen helppokäyttöinen ja tehokas. Lisäksi käyttäjän ei tarvitsisi olla perehtynyt ontologisoidun tiedon aihealueeseen ymmärtääkseen haun muodostusta. (Lei, et al., 2006)

Lei et. al. ovat kuitenkin tutkimuksessaan osoittaneet vuonna 2006, että semanttisen webin teknologioihin perustuvalla hakukoneella voidaan tuottaa hakukäyttöliittymiä, jossa käyttäjälle luodaan yksinkertaisia näkymiä, joista voidaan suorittaa tehokkaita hakuja. Vastaavasti käyttäjälle esitetyt hakutulokset tyydyttävät hakutarpeet ja ne mm. selittävät ”itsensä auki”, jolloin käyttäjä ymmärtää esimerkiksi miksi on juuri tämän hakutuloksen saanut. (Lei, et al., 2006)

Kun nykypäivänä puhutaan tiedon hakemisesta, ei voida sivuuttaa Googlea tai Microsoftin Bing hakukonetta. Niin Googella kuin Microsoftilla on omat semanttisen webin ja ontologioihin pohjautuvat hakunsa. Googlen hakuja parannetaan heidän omalla The Knowledge Graphilla, jonka avulla sanoilla annetaan merkitys ja tämän avulla hakijalle voidaan tarjota lisätietoa. Vastaavasti Microsoftin Satori suorittaa samaan tehtävää. Molemmissa ideana on määrittää entiteettejä, niiden ominaisuuksia sekä näiden välisiä suhteita. (Farber, 2013) Kyseessä on siis sama idea kuin avoimessa linkitetyssä datassa. Kuvassa 10 on Google hausta kuvakaappaus, jossa voidaan nähdä yksi esimerkki Googlen The Knowledge Graphin toiminnasta.

matti nykänen

Verkkohaku Kuvahaku Videot Kartat Sovellukset Lisää Hakutyökalut

Noin 393 000 tulosta (0,25 sekuntia)

Nykänen Matti Keikat - magnumlive.fi
www.magnumlive.fi/matti-nykanen/
 Tee tapahtumastasi menestys - Matti vetää salin täyteen. Pyydä tarjous!

Matti Nykänen – Wikipedia
fi.wikipedia.org/wiki/Matti_Nykänen
 Matti Ensio Nykänen (ent. Paanala (1996–1998), s. 17. heinäkuuta 1963 Jyväskylä) on suomalainen nelinkertainen mäkihypyn olympiavoittaja. Nykänen voitti ...
 Urheilu-ura - Julkisuudessa - Yksityiselämä - Saavutuksia ja palkintoja

Nykänen Matti - Julkkikset - SeiskaLive - 7 Päivää
www.seiska.fi/julkikkiset
 Matti Nykänen valmistautui vapun viettoon juomavarastojen huolellisella täyttämällä. ... Matti Nykäsen pukeutumistyyli järkyttää - persposikin vilkkuu! Kuvat!

Matti Nykänen tunnusti: Ratkesin juomaan, tuli ... - Iltalehti
www.iltalehti.fi/viihde/2015041219506637_vi.shtml
 12.4.2015 - Matti Nykänen kertoi 100 päivää ilman viinaa -finaalilähetyksessä ratkenneensa juomaan kesken haasteen.

Tältä näyttää dokumentti, jota Matti Nykänen vastustaa ...
www.hs.fi/kulttuuri/a1430102572143
 27.4.2015 - Matti Nykäsestä kertovasta dokumenttielokuvasta puhkesi viime perjantaina julkinen riita, kun mestarimäkihyppääjä Nykänen kertoi Helsingin ...

Matti Nykänen – Wikisitaatit
fi.wikiquote.org/wiki/Matti_Nykänen
 (28.10.2008 Nykänen seiskalle saamistaan jättisakoista); "Minäkin olen jonkinlainen mestari, otetaan erä". (Mervin mukaan Matin haasteessa karatemestarin ...

Matin elämäntapa - Matti Nykäsen alastulomonttuun
www.mattinykanen.net/thoughts.html
 Matti Nykäsen alastulomonttu! ... Tervetuloa Matti Nykäsen alastulomonttuun! ...
 Tuntojemme tulkin, avaruuden anturin - Matin - korkealentoisia pohdintoja ...

Matti Nykänen
www.mattinykanen.net/
 Matti-elokuvan arvio lisätty uuteen NykäLeffa-osioon. Eikä menny ku ... Nelosen 4D esittää "tosity-dokumentin": Matti Nykänen - Uuteen nousuun! 28.02.08

Matti Nykänen
 Olympiavoittaja






Matti Ensio Nykänen on suomalainen nelinkertainen mäkihypyn olympiavoittaja. Nykänen voitti urallaan viisi olympiamitalia, 14 maailmanmestaruusmitalia ja 22 Suomen-mestaruusmitalia. [Wikipedia](#)

Pituus: 1,77 m
 Puolisot: Mervi Tapola (v. 2004), Mervi Tapola (v. 2001–2003), lisää
 Lapset: Sami Nykänen, Eveliina Nykänen

Kappaleet

Elämä on laiffii	2006	Ehkä otin, ehkä en
Ehkä otin, ehkä en	2006	Ehkä otin, ehkä en
Jokainen tsäänsi on m...		
V-tyyli	1992	Yllätysten yö
Mäkiotko ja häkiliintu	2006	Ehkä otin, ehkä en

Käyttäjät hakivat myös näitä Näytä 15+ lisää

				
Jens Weißflog	Janne Ahonen	Gregor Schlieren...	Harri Olli	Matti Hautamäki

Kuva 10. Googlen hakutulos, jossa rikastettua tietoa The Knowledge Graphin avulla (<https://www.google.fi/search?q=matti+nyk%C3%A4nen&ei=CuVqVfn6JsRsAHv2IGwDg#>)

Kuvan oikeassa reunassa on esitetty Googlen The Knowledge Graphin avulla rikastettua tietoa entiteetistä ”Matti Nykänen”. Tiedoista käy ilmi esimerkiksi ”Pituus”, ”Kappaleet” sekä muiden ”Matti Nykästä” hakeneiden käyttäjien hakemia entiteettejä (tässä yhteydessä todennäköisesti ”henkilö” entiteettejä).

Lisäksi esimerkiksi Google The Knowledge Graph toimii Googlen Now-palvelun taustalla. Now-palvelu on uusimmissa Android-puhelimissa tarjottu sovellus, joka toimii käyttäjän ”henkilökohtaisena apurina”. Esimerkiksi sovellukselta voisi suoraan kysyä ”Kuinka pitkä on Matti Nykänen” (englanniksi), jolloin palvelu osaisi vastata The Knowledge Graphista löytyvän tiedon pohjalta. (Bohn, 2012)

Vaikka luvun alussa todettiin nykyisen WWW:n olevan puutteellinen hakemisen osalta perinteisessä internetissä, on tiedonhaku ns. suurten toimijoiden kuten Googlen ja Microsoftinkin tekemän kehityksen myötä kehittynyt hyödyntämällä semanttisen webin

teknologioita. Tämän pohjalta voidaan todeta, että semanttisen webin teknologioiden, ontologioiden sekä linkitetyn datan avulla voidaan tarjota käyttäjille parempia hakuja. Muuten ei olisi perusteltua, että em. voittoa tavoittelevat yritykset panostaisivat näin laajamittaisesti semanttisen webin teknologioihin.

4.2.2. Ontologisoidun paikkatiedon hyödyntäminen hakemisessa

Paikkatietoa voidaan hyödyntää hakemisessa esimerkiksi tarjoamalla hakijalle hakutuloksiksi määritellyn paikkakohteen ”vieressä” olevia kohteita. Tätä mallia on testattu esimerkiksi GeoSharen (<http://web.ics.purdue.edu/~liu207/Geoshare%20-%20Home.htm>) puolesta, jossa pyrittiin luomaan käyttäjälle reitti ”puhtaan veden” äärellä tietyssä sijainnissa. He käyttivät termiä ”spatial relevance reasoning” tilanteessa, jossa käyttäjälle voitiin suositella paikkoja käyttäjän määrittämän paikan perusteella. Ehdotetut paikat perustuivat etäisyyksiin alkuperäisestä paikasta. (Hübner, et al., 2007)

Fu et. al. niin ikään on osoittanut, miten ontologisoitua spatiaalista tietoa voidaan hyödyntää dokumenttien löytämisessä ja hakemisessa. Ratkaisu perustuu sumeisiin paikkatieto suhteisiin (fuzzy spatial relationship) ja näistä johdettuihin suosituksiin. Käyttäjälle pystytään tällöin tarjoamaan merkittävästi parempia hakuja. (Fu, et al., 2005) Merkittävän arkistojen näkökulmasta tästä tutkimuksesta tekee se, että tutkimuksessa on hyödynnetty dokumentteihin liittyviä paikkatietoja, sekä se, että paikkatieto-ontologiat ovat olleet ”reaalimaailmasta” eivätkä vain keinotekoisesti luotuja.

4.3. SAPO

SAPO eli Suomen ajallinen paikkaontologia sisältää tietoa paikoista, joille voidaan antaa ajallinen kattavuus. Nykyisin SAPOn instanssit eli kohteet koostuvat Suomen kuntahistorian ajan kunnista sekä näiden ajallisista kattavuuksista eli alkua- ja päättymispäivistä. Lisäksi kunnista on osittain kerrottu niiden koordinaattitietoa eli kuntien rajat wgs84 polygoneina sekä muutamia muita tietoja kuten esimerkiksi muunkieliset nimet ja tyypit. (Finto, 2014)

Toistaiseksi SAPOa ei ole julkaistu tuotantoon, sillä sen ylläpidosta ei ole saatu sovittua kansallisesti. Tästä syystä SAPO on julkaistu vain Finto-portaalin kehityspuolella ja tämän työn ratkaisussa hyödynnetään SAPOn kehitysversiona.

SAPO soveltuu arkistolaitoksen ja arkistojen näkökulmasta hyvin testattavaksi kohteeksi, sillä arkistot ovat joutuneet toimintaympäristön pakottamina hallinnoimaan aineistojaan ja niiden kuvailutietoja kokonaisuuksina. Tämä on aiheuttanut sen, että myös esimerkiksi paikannimiä on tuotettu enemmän yleisemmällä tasolla. Eli ei olla esimerkiksi tuotettu tietoja kylistä vaan ennemminkin kunnista. Toisaalta digitoinnin myötä tarve tarkemmalle metatietojen tuottamiselle on kasvanut.

SAPOn potentiaalin haun taustalla toimivana ontologiana perustuu sen sisältämään yhdistelmätietoon ajallisesta ja paikallisesta kattavuudesta. Kuntapolygonien ja näiden ajallisten kattavuuksien avulla käyttäjälle voidaan suositella hakutuloksia viereisistä kunnista niin ajallisesti kuina paikallisestikin. Hakuja voidaan laajentaa ja rajata vastavasti. Vielä laajempaa hyötyä SAPOsta olisi, mikäli tätä laajennettaisiin hierarkiassa kunta-tasolta ylös- ja alaspäin siten, että esimerkiksi läänit eri aikoina otettaisiin mukaan tai tarkempia kuin kuntatasoisia instansseja voitaisiin kiinnittää nykyisten SAPOn kuntainstanssien alle.

Vaikka SAPO tietyssä mielessä sopii todennäköisesti arkistoaineistolle kohtalaisen hyvin, ei se riitä kuitenkaan ainoaksi paikkatietolähteeksi. Tarpeet kohdistuvat paikannimien ja paikkatiedon osalta myös epätarkempia ja tarkempien kohteiden määrittelyyn. Lisäksi SAPO sisältää vain Suomen kunnat, kun taas tarpeet ovat käytännössä globaalit.

5. Ratkaisun vaatimukset ja ratkaisun kuvaus

Ratkaisun tavoitteena on selvittää, miten hyvin nykyinen paikannimistö arkistolaitoksen aineistojen metatiedoista saadaan linkitettyä Finton tarjoamaan avoimeen linkitettyyn dataan SAPO-ontologian osalta. Ratkaisu perustuu vallitsevaan nykytilanteeseen, tavoitetaan sekä olemassa olevan teknologian hyödyntämiseen. Ratkaisusta on kuvattu lähtötilanne sekä vaatimukset toteutettavalle ratkaisulle. Ratkaisun on toteuttanut arkistolaitoksen suunnittelija tässä työssä määriteltyjen vaatimusten pohjalta ja vaatimuksia on päivitetty keskustelujen ja tarkennusten pohjalta. Toteutuksen kuvaus perustuu yksinomaan suunnittelijan kertomaan (luku 5.4).

Ratkaisussa on pyritty tuomaan esille se, kuinka paljon paikannimistöä arkistojen kuvailumetatiedoissa nykyisin on käytetty sekä se, miten hyvin nämä paikannimet voitaisiin linkittää SAPOn entiteetteihin huomioimalla ajallinen kattavuus. Lopputuloksena on syntynyt arvio paikannimien käytöstä ja niiden ”osumistarkkuudesta” SAPOon. Tulokset ja näiden arviointi on käsitelty luvussa 6.

5.1. Valittu aineisto

Arkistolaitoksen aineistojen metatietoja käsitellään nykyisin kolmessa eri järjestelmässä, kuten luvussa 2 on todettu. Koska Vakka-arkistotietokanta on laajin ja kattavin arkistolaitoksen aineistojen metatietoa sisältävistä järjestelmistä, on sen sisältämä kuvailutieto valittu tämän työn aineistoksi. Vakka-arkistotietokanta sisältää metatiedot lähes kaikista arkistolaitoksen kuvailemista arkistonmuodostajista ja aineistoista. Vakka-arkistotietokannan täydentämisessä käytetty yhteisiä kuvailu- ja luettelointisääntöjä, joten teoriassa syötettyjen tietojen tulisi olla yhteismitallisia. Vakka-arkistotietokannassa arkistokuvailun metatietoja hallinnoidaan hierarkisesti, kuten kuvassa 4 on esitetty. Kuitenkin vain kahdella hierarkian tasolla; arkistonmuodostajilla ja arkistoyksiköillä on osoitettu omat kenttensä paikannimitiedon hallinnointiin. Ratkaisuun valittu aineisto koostuu siten Vakka-arkistotietokannan arkistonmuodostajien ja arkistoyksiköiden metatiedoista löydetystä paikannimistä sekä näiden ajallisista kattavuuksista. Ajalliset kattavuudet ovat merkittäviä työn kannalta, sillä Suomen ajallinen paikkaontologia tunnistaa eri aikakausien poikkeavat kunnat omina entiteetteinään.

5.2. Arviointimenetelmä

Tutkimuskysymykseen: ”Miten hyvin suomalainen ajallinen paikkaontologia soveltuu arkistolaitoksen paikannimitiedon rikastamiseen?” pyritään vastaamaan muodostamalla käsitys siitä, miten hyvin paikannimitietoa on löydettävissä arkistonmuodostajilta ja arkistoyksiköiltä sekä miten hyvin löydetyille paikannimille löydetään vastineet SA-POsta. Molemmissa arvioinneissa käytetään määrällistä analyysiä

(<https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/aineiston-analyysimenetelmat/maarallinen-analyysi>), mutta työn pohdinnassa käytetään myös laadullista analyysiä

(<https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/aineiston-analyysimenetelmat/laadullinen-analyysi>). Ratkaisussa analysoidaan vertailulukuja, jotka saadaan laskemalla osumatarkkuuksia prosentein. Toisaalta myös tuotettua dataa analysoidaan sisällöllisesti ja pyritään tuomaan esiin, mitkä asiat vaikuttavat tuloksen muodostumiseen.

5.3. Vaatimukset toteutukselle

Lähtökohtana toteutuksella voidaan pitää tilannetta, jossa nykytilassa arkistojen meta-tiedoissa kuvailukohteilta löytyy paikannimistöä ja ajallista tietoa sekä SAPOsta löytyy paikannimistöä, jolla on ajallinen kattavuus. Arkistojen paikannimistö koostuu tekstistä, joka joudutaan automatisoidusti poimimaan sen suuren määrän vuoksi. Paikannimien onnistunut poiminta on edellytys onnistuneelle ratkaisulle. Lisäksi kuvailukohteissa, joissa paikannimistöä esiintyy, voi löytyä myös ajallista tietoa. Aikatiedot on nykyisin merkitty alkuvuodella ja loppuvuodella, jotka kuvaavat kuvailukohteenkohteen ”ajallista kattavuutta”. Ajallinen kattavuus on merkitty vuoden, vuosikymmenen, vuosisadan tai vuosituhannen tarkkuudella. Johtuen käytänteistä aikatietoa ei aina kuitenkaan löydy ns. täydellisesti eli siten, että alkua- ja loppuaikat olisivat molemmat määritetty. On myös mahdollista, että aikatietoa ei löydy lainkaan. Tällöin seuraavat vaihtoehtoiset tilanteet kuvailukohteiden ajallisen kattavuuden määrittämisessä ovat mahdollisia:

1. Alku- ja loppuaika tiedetään esim. 1900-1910
2. Vain alkuaika tiedetään esim. 1900-
3. Vain loppuaika tiedetään esim. -1910
4. Alku- ja loppuaikaa ei tiedetä

SAPOssa taas on tilanne, jossa alkuaika on aina tiedossa, mutta loppuaika ei välttämättä. Tällöin loppuajan poisjäänti tarkoittaa sitä, että kohde on nykyisin voimassa oleva eli nykyinen kunta. Tällöin meillä on kaksi vaihtoehtoista tilannetta:

1. Alku- ja loppuaika tiedetään esim. 1900-1910
2. Vain alkuaika tiedetään esim. 1900-

Lisäksi kohdeontologiassa on kohteita, joilla ei ole aikatietoa ja nämä ovat aina aikatie-toa sisältävien kohteiden yläkäsitteitä.

Toteutuksessa tulee ensin selvittää mitkä lähtötiedoissa esiintyvät paikannimet löydetään SAPOsta. Tämän jälkeen voidaan selvittää, mitkä lähtötiedoissa olevat paikannimet liittyvät johonkin tietyn aikakauden paikkaa eli tiettyyn SAPOn aikapaikka – käsitteeseen. Näiden tietojen pohjalta voimme rakentaa seuraavat vaatimukset tietojen poiminnalle:

Otetaan arkistokuvailusta esimerkiksi kohteet, joista löytyy paikannimenä määritelty ”Kaarina”, mutta eriäviä aikamääriytyksiä. Aikamääriytykset eroavat edellä esitettyjen vaihtoehtojen mukaisesti.

Otetaan esimerkkikohteeksi SAPOsta Kaarina ja tämän alakäsitteet, joille on määritelty ajalliset kattavuudet:

Kaarina, URI: <http://www.yso.fi/onto/sapo/Kaarina>

Kaarina (1869-01-01), ajallinen kattavuus 1.1.1869-31.12.1935, URI:
http://www.yso.fi/onto/sapo/Kaarina_1869_01_01

Kaarina (1936-01-01), ajallinen kattavuus 1.1.1936-31.12.1938, URI:
http://www.yso.fi/onto/sapo/Kaarina_1936_01_01

Kaarina (1939-01-01), ajallinen kattavuus 1.1.1939-31.12.1945, URI:
http://www.yso.fi/onto/sapo/Kaarina_1939_01_01

Kaarina (1946-01-01), ajallinen kattavuus 1.1.1946-31.12.2008, URI:

http://www.yso.fi/onto/sapo/Kaarina_1946_01_01

Kaarina (2009-01-01), ajallinen kattavuus 1.1.2009-, URI:

http://www.yso.fi/onto/sapo/Kaarina_2009_01_01

Tämän jälkeen toimitaan seuraavien sääntöjen mukaan. Aikojen osalta tarkin käsittely-taso on vuosittain, sillä arkistojen metatiedoissa tarkempia tietoja ei ole käytetty:

1. Jos lähtötiedot ovat täydelliset eli alku- ja loppuaika tiedetään
 - a. Poimitaan SAPOsta kaikki ne kohteet, joissa löydetyn kohteen ajallinen kattavuus on päällekkäinen alkuperäisen kohteen ajallisen kattavuuden kanssa
 - i. Esim. Arkistojen kuvailukohteesta on löydetty termi Kaarina ja tämän ajallinen kattavuus 1937-1960. Saposta löytyy kohteet Kaarina (1869-1935) Kaarina (1936-1938), Kaarina (1939-1945), Kaarina (1946-2008) ja Kaarina (2009-). Tällöin poimittaisiin kohteet Kaarina (1936-1938), Kaarina (1939-1945) ja Kaarina (1946-2008).
2. Jos lähtötiedot ovat puutteelliset loppuvuoden osalta
 - a. Poimitaan SAPOsta kaikki ne kohteet, jotka osuvat osittain tai kokonaan alkuvuoden jälkeiselle ajanjaksolle.
 - i. Esim. Arkistojen kuvailukohteesta on löydetty termi Kaarina ja tämän ajallinen kattavuus on 1950-. Tällöin poimittaisiin kohteet Kaarina (1946-2008) ja Kaarina (2009-).
3. Jos lähtötiedot ovat puutteelliset alkuvuoden osalta
 - a. Poimitaan SAPOsta kaikki ne kohteet, jotka osuvat osittain tai kokonaan loppuvuotta edeltävälle ajanjaksolle.
 - i. Esim. Arkistojen kuvailukohteesta on löydetty termi Kaarina ja tämän ajallinen kattavuus on -1950. Tällöin poimittaisiin kohteet Kaarina (1869-1935) Kaarina (1936-1938), Kaarina (1939-1945) ja Kaarina (1946-2008).
4. Jos lähtötiedot ovat puutteelliset ajan osalta kokonaan
 - a. Poimitaan SAPOsta vain yläkäsite, jolla ei ole ajallista kattavuutta.

- i. Esim. Arkistojen kuvailukohteesta on löydetty termi Kaarina ja tälle ei ole määritelty ajallista kattavuutta. Tällöin poimittaisiin vain kohde Kaarina (yläkäsité).

Koska halutaan analysoida myös SAPOn soveltuvuutta nykyisille metatiedoille, tulee toteutuksessa myös tuottaa tieto siitä, mille kuvailtaville kohteille, joista paikannimistöä löytyi, löytyi vähintään yksi vastine SAPOsta. Lisäksi tulee toteutuksen lopputuloksesta käydä ilmi mille kohteille ja millaisille paikannimille ei löytynyt vastinetta SAPOsta.

Toisin sanoen tarvitsemme listauksen kuvailukohteista:

1. Joista löytyy paikannimi tai paikannimiä
2. Näistä kohteista paikannimi tai paikannimet, jotka löydettiin
3. Näiden kohteiden osalta tieto, löydettiinkö SAPOsta vastinetta yhdelle tai useammalle paikannimelle, jota kuvailukohteessa oli käytetty
 - a. Ja näiden kohteiden osalta listattuna kohteet SAPOsta, jotka poimittiin

Tämän lisäksi tarvitaan tieto siitä, kuinka paljon kuvailukohteita kokonaisuudessaan on, jotta voimme arvioida, kuinka paljon paikannimitietoa on tuotettuna.

5.4. Toteutuksen kuvaus

Vakka-arkistotietokannan ja AHAA:n välille on tuotettu migraatiotyökalu, joka poimii arkisto kerallaan kokonaisuuksia Vakasta. Migraatiotyökalussa on työnkulku, jonka sisällä kuvailukohteiden metatiedoista poimitaan maantieteellistä kattavuutta kuvaavista kentistä paikannimiä arkistonmuodostaja ja arkistoyksikkö tasoilta kuten luvussa 2.3 on todettu (kts. myös Kuva 4). Paikannimet on poimittu kahdesta eri kentästä; PAIKKA-KUNTA (myöh. Paikan nimi) ja MUU_ALUE (myöh. Muu paikka). Työnkulussa näistä paikannimistä generoidaan kutsuja, joita hyödynnetään kutsuessa paikannimiontologian rajapintoja. Vastauksena rajapinta palauttaa mm. kohteen URI:n (Uniform Resource Identifier), jonka avulla kohteesta voidaan pyytää lisää tietoa. Lisätietoina pyydetään kohteen olemassaolon alkuaika ja päättymisaika eli ”existenceBeginsAt” ja ”existenceEndsAt”. Tämän jälkeen työnkulku vertailee kyseessä olevan kuvailukohteen ajallista kattavuutta SAPO kohteen aika-arvoihin. Näiden tietojen pohjalta tuotetaan analyysitiedosto CSV-muodossa, joka voidaan viedä excel-tiedostoon. (Laukkonen, 2015) Tiedostoon tulostuu tiedot seuraavasti (A,B,C,D,E ovat excel tiedoston sarakkeet):

A=AM_TUNNUS,

B=Paikan nimi (AM_ALKUVUOSI,AM_LOPPUVUOSI)

C=Paikkojen URI

D=Muu paikka (AM_ALKUVUOSI,AM_LOPPUVUOSI)

E=Muiden paikkojen uri:t (Laukkonen, 2015)

AM_TUNNUS tarkoittaa tässä yhteydessä arkistonmuodostaja kuvailukohteen Vakka-arkistotietokannan sisäistä teknistä tunnusta. Vastaavasti arkistoyksiköiden osalta tuotetaan vastaava excel-tiedosto, jossa ”AM_” korvautuu ”AY_”:llä, koska käytetään arkistoyksikön tietoja.

Tiedostot halutaan erottaa arkistonmuodostajien ja arkistoyksiköiden osalta sen takia, että kyseessä on suuri datamassa ja näiden käsittely yhdessä voisi olla haastavaa. Lisäksi työn kannalta on merkityksellistä tehdä huomioita siitä, miten kahdella hyvin eri tarkkuustasoisella kuvailukohteella, arkistonmuodostajalla ja arkistoyksiköllä, paikannimistää on käytetty (kts. Kuva 4).

6. Tulokset

Ratkaisun tuloksena syntyi kaksi excel-dokumenttia, joista toinen sisältää tietoa arkistonmuodostajista poimittujen paikannimien ja aikojen perusteella löydettyjä kohteita SAPOsta ja vastaavasti toinen sisältää samat tiedot, mutta arkistoyksikkötasolta. Seuraavissa alaluvuissa on esitetty ratkaisun tulokset.

Excel dokumentit sisälsivät luvussa 5.4 kuvatut excel-taulukot, joiden ulkomuoto on esitetty kuvassa 11. Kuvasta huomataan esimerkiksi riviltä 5241, että kohteeseen ”Turku (1964-1996)” sarakkeessa ”Paikannimi” voidaan liittää neljä eri SAPOsta löytyvää ajallista kohdetta koskien Turku. Nämä ovat

http://www.yso.fi/onto/sapo/Turku_1968_01_01,

http://www.yso.fi/onto/sapo/Turku_1967_01_01,

http://www.yso.fi/onto/sapo/Turku_1973_01_01 ja

http://www.yso.fi/onto/sapo/Turku_1949_01_01. Vastaavasti rivillä 5240 kentässä

”Muu paikka” sarakkeessa olevaan ”Kivennapa (1917-2006)” kohteeseen voidaan liittää

SAPOsta kohteet http://www.yso.fi/onto/sapo/Kivennapa_1937_01_01 ja

http://www.yso.fi/onto/sapo/Kivennapa_1911_01_01.

	A	B	C	D	E
	AM_TUNNUS	Paikannimi (AM-ALKUVUOSI-AMLOPPUVUOSI)	Paikkojen URI:t	Muu paikka (AM-ALKUVUOSI,AM-LOPPUVUOSI)	Muiden paikkojen URI:t
1		AM_TUNNUS: (Kaikki näkyvät)	http://www.yso.fi/onto/sapo/Turku_1290_01_01 , http://www.yso.fi/onto/sapo/Turku_1968_01_01 , http://www.yso.fi/onto/sapo/Turku_1967_01_01 , http://www.yso.fi/onto/sapo/Turku_1944_01_01 , http://www.yso.fi/onto/sapo/Turku_1939_01_01 , http://www.yso.fi/onto/sapo/Turku_1930_01_01 , http://www.yso.fi/onto/sapo/Turku_1973_01_01 , http://www.yso.fi/onto/sapo/Turku_1949_01_01 ,		
5240	38668.KA	Turku (1917-2006)		Kivennapa (1917-2006)	http://www.yso.fi/onto/sapo/Kivennapa_1937_01_01 , http://www.yso.fi/onto/sapo/Kivennapa_1911_01_01 ,
			http://www.yso.fi/onto/sapo/Turku_1968_01_01 , http://www.yso.fi/onto/sapo/Turku_1967_01_01 , http://www.yso.fi/onto/sapo/Turku_1973_01_01 , http://www.yso.fi/onto/sapo/Turku_1949_01_01 ,		
5241	38670.KA	Turku (1964-1996)			
			http://www.yso.fi/onto/sapo/Oulu_2013_01_01 , http://www.yso.fi/onto/sapo/Oulu_1965_01_01 , http://www.yso.fi/onto/sapo/Oulu_1938_01_01 , http://www.yso.fi/onto/sapo/Oulu_1961_01_01 , http://www.yso.fi/onto/sapo/Oulu_1947_01_01 , http://www.yso.fi/onto/sapo/Oulu_1886_01_01 , http://www.yso.fi/onto/sapo/Oulu_1605_01_01 , http://www.yso.fi/onto/sapo/Oulu_1979_01_01 , http://www.yso.fi/onto/sapo/Oulu_2009_01_01 , http://www.yso.fi/onto/sapo/Oulu_1911_01_01 ,		
5242	38671.KA	Oulu (-)			

Kuva 11. Esimerkkiotos AM excel-taulukosta. (Laukkonen, 2015)

Tulokset on tuotettu excel-taulukoista tietoja poimimalla ja laskemalla. Näiden poimintojen ja laskemisen pohjalta on tuotettu seuraavissa alaluvuissa esitetyt analysointitaulukot. Tulosten analysoinnissa on käytetty excel-ohjelmaa.

6.1. Arkistonmuodostaja -kuvailukohteista löydetty paikannimistö ja SAPO

Arkistomuodostajien osalta taulukossa 3 on esitetty kuinka paljon arkistomuodostaja -kuvailukohteita on yhteensä sekä näistä löydetty kohteet, jolle on merkitty paikkatietoa jompaankumpaan em. mainittuun kenttään ("Paikkakunta", "Muu paikka"). Määrät ovat kentittäin eriteltynä. Lisäksi taulukkoon on laskettu näiden prosentuaalinen osuus kaikista arkistonmuodostajista eli esimerkiksi 7953 arkistonmuodostajalta löytyi paikkatietoa ja tämä osuus on 24,7 % kaikista arkistonmuodostajista. Lisäksi taulukossa on kenttien käyttöprosentti, kun tarkastellaan kumpaa kenttää on käytetty, jos paikkatietoa on merkitty. Eli esimerkiksi "Paikkakunta" –kenttää on käytetty 92,1 % arkistonmuodostajista, joille on määritetty jotain paikkatietoa.

Taulukko 3. Arkistonmuodostajien määrä ja näistä löydetty paikannimistö.

	Arkistonmuodostajia yhteensä	Arkistonmuodostajia, joista löytyi paikkatietoa	Arkistonmuodostajia, joista löytyi tietoa kentästä "Paikkakunta"	Arkistonmuodostajia, joista löytyi tietoa kentästä "Muu paikka"
Määrä	32154	7953	7328	1218
Prosenttiosuus kaikista	100,0 %	24,7 %	22,8 %	3,8 %
Prosenttiosuus paikkatiedoista			92,1 %	15,3 %

Taulukosta 3 huomataan, että paikannimistöä on merkitty noin neljäsosalle kaikista arkistonmuodostajista (24,7 %). Lisäksi voidaan todeta, että "Paikkakunta" –kenttää on käytetty huomattavasti systemaattisemmin kuin "Muu paikka" kenttää. Vertailemalla prosenttiosuutta paikkatiedoista, huomaamme, että "Paikkakunta" ja "Muu paikka" –kenttien prosenttiosuuksien summa ylittää 100 %. Tämä osoittaa sen, että 7,4 % (92,1 % + 15,3 % - 100 %) kohteista, jossa paikkatietoa on määritetty, on sitä määritetty molempiin kenttiin.

Taulukossa 4 on esitetty tiedot siitä, kuinka monelle kuvailukohteen paikannimelle on löydetty vähintään yksi vastine SAPOsta sekä tieto siitä, kuinka suurelle prosenttiosuudelle se on löydetty kun verrataan kentän käyttöön.

Taulukko 4. Arkistonmuodostajien paikkatietojen löytyminen SAPOsta.

	Arkistonmuodostajia yhteensä	Arkistonmuodostajia, joista löytyi tietoa kentästä "Paikkakunta"	Arkistonmuodostajia, joiden "Paikkakunta" kentälle löydettiin vähintään yksi vastine SAPOsta	Arkistonmuodostajia, joista löytyi tietoa kentästä "Muu paikka"	Arkistonmuodostajia, joiden "Muu paikka" kentälle löydettiin vähintään yksi vastine SAPOsta
Määrä	32154	7328	6564	1218	131
Prosenttiosuus kaikista arkistonmuodostajista	100,0 %	22,8 %	20,4 %	3,8 %	0,4 %
Prosenttiosuus jolle löytyi vastine SAPOsta kun verrataan käytettyyn kenttään			89,6 %		10,8 %

Taulukosta 4 huomataan, että vastine on löytynyt "Paikkakunta" –kenttää käytettäessä lähes kaikista paikannimistä, joita on käytetty (89,6 %). "Muu paikka" –kenttää käytettäessä vastinetta ei kuitenkaan löytynyt kovinkaan suuressa osassa, vaan prosentti jää

hyvin pieneksi (10,8 %). Lisäksi huomataan, että SAPO vastine vähintään 82,5 % ((6564)/7953) kaikista arkistonmuodostajakohteista, joille paikannimitietoa on merkitty. Luku on ”vähintään”, sillä tarkkaa arvoa on haastava laskea, sillä osassa arkistonmuodostajista paikkatietoa on syötetty molempiin ”Paikkakunta” ja ”Muu alue” –kenttiin. Vastaavasti huomataan, että SAPO vastine kaikille arkistonmuodostajille löytyy vähintään 20,4 % kohteista.

6.2. Arkistoyksikkö kuvailukohteista löydetty paikannimistö ja SAPO

Arkistoyksiköiden osalta taulukossa 5 on esitetty kuinka paljon arkistoyksikkö kuvailukohteita on yhteensä sekä näistä löydetty kohteet, jolle on merkitty paikkatietoa jompaankumpaan em. mainittuun kenttään sekä määrät kentittäin eriteltynä. Lisäksi taulukoon on laskettu näiden prosentuaalinen osuus kaikista arkistoyksiköistä. Lisäksi taulukossa on kenttien käyttöprosentti, kun tarkastellaan kumpaa kenttää on käytetty, jos paikkatietoa on merkitty.

Taulukko 5. Arkistoyksiköiden määrä ja näistä löydetty paikannimistö.

	Arkistoyksiköjä yhteensä	Arkistoyksiköjä, joista löytyi paikkatietoa	Arkistoyksiköjä, joista löytyi tietoa kentästä "Paikkakunta"	Arkistoyksiköjä, joista löytyi tietoa kentästä "Muu paikka"
Määrä	3347850	388115	374411	220463
Prosenttiosuus kaikista	100,0 %	11,6 %	11,2 %	6,6 %
Prosenttiosuus paikkatiedoista			96,5 %	56,8 %

Taulukosta 5 huomataan, että paikannimistöä on merkitty vain hieman yli 10 % kaikista arkistoyksiköistä (11,6 %). Lisäksi voidaan todeta, että ”Paikkakunta” –kenttää on käytetty systemaattisemmin kuin ”Muu paikka” kenttää (11,2 % vs. 6,6 %). Vertailemalla prosenttiosuutta paikkatiedoista, huomaamme, että ”Paikkakunta” ja ”Muu paikka” –kenttien prosenttiosuuksien summa ylittää 100 %. Tämä osoittaa sen, että 53,3 % (96,5 % + 56,8 % - 100 %) kohteista, jossa paikkatietoa on määritelty, on sitä määritelty molempiin kenttiin.

Taulukko 6 esittää tiedot siitä, kuinka monelle kuvailukohteen paikannimelle on löydetty vähintään yksi vastine SAPOsta sekä tieto siitä, kuinka suurelle prosenttiosuudelle se on löydetty kun verrataan kentän käyttöön.

Taulukko 6. Arkistoyksiköiden paikkatietojen löytyminen SAPOsta.

	Arkistoyksiköjä yhteensä	Arkistoyksiköjä, joista löytyi tietoa kentästä "Paikkakunta"	Arkistoyksiköjä, joiden "Paikkakunta" kentälle löydettiin vähintään yksi vastine SAPOsta	Arkistoyksiköjä, joista löytyi tietoa kentästä "Muu paikka"	Arkistoyksiköjä, joiden "Muu paikka" kentälle löydettiin vähintään yksi vastine SAPOsta
Määrä	3347850	374411	265118	220463	22183
Prosenttiosuus kaikista arkistoyksiköistä	100,0 %	11,2 %	7,9 %	6,6 %	0,7 %
Prosenttiosuus jolle löytyi vastine SAPOsta kun verrataan käytettyyn kenttään			70,8 %		10,1 %

Taulukosta 6 huomataan, että vastine on löytynyt ”Paikkakunta” –kenttää käytettäessä huomattavasti yli puolesta kohteista (70,8 %). ”Muu paikka” –kenttää käytettäessä vastinetta ei kuitenkaan löytynyt kovinkaan suuressa osassa, vaan prosentti jää hyvin pieneksi (10,1 %). Lisäksi huomataan, että SAPO vastine löytyy vähintään 68,3 %

((265118/388115)) kaikista arkistoyksikkökohteista, joille paikannimitietoa on merkitty. Luku on ”vähintään”, sillä tarkkaa arvoa on haastava laskea, sillä osassa arkistoyksiköistä paikkatietoa on syötetty molempiin ”Paikkakunta” ja ”Muu alue” –kenttiin. Vastaavasti huomataan, että SAPO vastine kaikille arkistoyksiköille löytyy vähintään 7,9 % kaikista arkistoyksiköistä.

6.3. Tulosten arviointi

Tuloksista voidaan huomata muutamia keskeisiä asioita. Yksi merkittävimmistä on se, että paikkatietoa tai paikannimitietoa on tuotettu systemaattisemmin arkistonmuodostajille kuin arkistoyksiköille (22,8 % vs. 11,2 %). Lisäksi nämä luvut osoittavat sen, että paikkatieto on koettu arkistojen luetteloinnissa merkittäväksi asiaksi, sillä kaikki tuotettu paikkatieto on ollut vapaaehtoista tuottaa ainakin käytettävän työkalun puolesta. Vakka-arkistotietokanta ei siis ole edellyttänyt paikkatiedon merkitsemistä kuvailukohteelle.

Toinen merkittävä asia on se, että huomattavasti enemmän käytettyä paikkatiedon tallennuskenttää (Paikkakunta) käytettäessä on vastine SAPOsta löytynyt huomattavin suurelle osuudella arkistonmuodostajia (89,6 %). Arkistoyksiköiden kohdallakin luku (70,8 %) on myös suuri. Mielenkiintoista on myös se, että ”Paikkakunta” kenttä sisälsi huomattavasti systemaattisemmin kuntatason tietoa kuin ”Muu alue” (89,6 % vs. 10,8 % ja 70,8 % vs. 10,1 %). Tämä voi olla ymmärrettävää, sillä kentän nimi on juuri ”Paikkakunta”. Erillistä ohjeistusta kuntatason tiedon merkitsemisestä ei kuitenkaan ole annettu.

Tulostaulukoista (excel-taulukot) voidaan kuitenkin huomata kaksi asiaa, jotka vaikuttavat tuloksiin ja niiden arviointiin. Vaikuttaisi siltä, että ”Muu alue” kenttää on käytetty usein silloin, kun paikaksi on merkitty epätarkempi kohde kuin kunta. Sitä on käytetty kuitenkin myös silloin kun ”Muu alue” tarkoittaa ”Paikkakuntaa”. Tämä huomio pätee arkistonmuodostajille ensisijaisesti. Taulukko 7 sisältää esimerkkiotoksen tästä.

Taulukko 7. Esimerkki ”Paikkakunta” ja ”Muu alue” kentistä löytyvistä tiedoista arkistonmuodostajilla.

Arkistonmuodostajan tekninen identifiointitunnus Vakka-arkistotietokannassa	”Paikkakunta” kentästä löytyvä merkkijono ja suluissa tieto ajallisesta kattavuudesta muodossa (”Alkuvuosi”-”Loppuvuosi”)	”Muu alue” kentästä löytyvä merkkijono ja suluissa tieto ajallisesta kattavuudesta muodossa (”Alkuvuosi”-”Loppuvuosi”)
38562.KA		Etelä-Pohjanmaa (-)
38563.KA		Keski-Pohjanmaa (-)
38566.KA	Utsjoki (1968-1986)	
38567.KA	Imatra (1999-2007)	
38568.KA	Turtola (1945-1952)	
38569.KA	Maaria (1911-1946)	Turku (1911-1946)
38584.KA	Jyväskylä (-)	Keljo (-)
38586.KA	Oulu (1967-)	
38587.KA	Laukaa (1946-)	
38590.KA	Janakkala (1937-2006)	
45242.KA	Rauma (-)	Lappi (-)
38594.KA	Maaria (1898-1952)	
3860.KA	Jyväskylä (1963-)	
38606.KA	Lappeenranta (-)	
38607.KA	Lappeenranta (-)	
38611.KA	Lappeenranta (1930-)	

38615.KA		Pohjois-Suomi (-)
38616.KA		Pohjois-Pohjanmaa (-)
38617.KA	Helsinki (1899-1900)	
39755.KA	Kouvola (1948-2006)	Myllykoski (1948-2006)

Kuten huomataan ”Keski-Pohjanmaa”, ”Pohjois-Suomi” ja ”Pohjois-Pohjanmaa” ovat epätarkempia ilmaisuja kuin kunnat ja nämä on merkitty ”Muu alue” –kenttään jättämällä ”Paikkakunta” kenttä samasta kohteesta tyhjäksi. Toisaalta taas ”Keljo” ja ”Myllykoski” tarkentavan ”Muu alue” –kentässä ilmaistuna saman kohteen ”Paikkakunta” kentässä kerrottua kuntaa.

Sama johtopäätös voidaan tehdä ainakin osittain arkistoyksiköiden osalta, mutta tämän luotettavuutta vähentää se, että kohteita on niin suuri määrä (>300 000). Näiden läpikäyminen perusteellisesti ei ole tehokasta, jolloin aineiston pinnallinen läpikäynti silmäillen on perusteltua. Yksistään näiden silmäily voi antaa kuitenkin varsin epäluotettavan kuvan. Silmiinpistävä huomio kuitenkin arkistoyksiköistä löydettyjen paikkatietojen osalta on se, että ne sisältävät kuitenkin selkeästi enemmän ulkomaisia paikkatietoja, jotka rajautuvat suoraan jo SAPOn ulkopuolelle. Taulukko 8 sisältää esimerkkiotoksen tästä.

Taulukko 8. Esimerkki ”Paikkakunta” ja ”Muu alue” kentistä löytyvistä tiedoista arkistoyksiköillä.

Arkistoyksikön tekninen identifiointitunnus Vakka-arkistotietokannassa	”Paikkakunta” kentästä löytyvä merkkijono ja suluissa tieto ajallisesta kattavuudesta muodossa (”Alkuvuosi”-”Loppuvuosi”)	”Muu alue” kentästä löytyvä merkkijono ja suluissa tieto ajallisesta kattavuudesta muodossa (”Alkuvuosi”-”Loppuvuosi”)
1324650.KA	Saksa (1954-1954)	Münster (1954-1954)
1324651.KA	Siilinjärvi (1961-1961)	
1324659.KA	Yhdysvallat (1911-1986)	Astoria (1911-1986)

1324671.KA	Taiwan (1974-1974)	
1324672.KA	Sveitsi (1958-1958)	Geneve (1958-1958)
1324675.KA	Patokoski, Australia (1957-1980)	
1324676.KA	Ylitornio (1986-1986)	
1324678.KA	Sarajärvi (1960-1960)	
1324682.KA	Helsinki (1955-1955)	
1324684.KA	Italia (1951-1951)	Rooma (1951-1951)

Kuten huomataan esiintyy ”Paikkakunta” –kentässä useita ulkomaisia maiden nimiä kuten ”Saksa”, ”Yhdysvallat”, ”Sveitsi” ja ”Italia”. Vastaavasti näiden ”Muu alue” kentissä on tarkennuksia näissä kohteissa kuten ”Münster”, ”Astoria”, ”Geneve” ja ”Rooma”.

7. Yhteenveto ja pohdintaa

Työssä pyrittiin vastamaan tutkimuskysymykseen: ”Miten hyvin suomalainen ajallinen paikkaontologia soveltuu arkistolaitoksen paikannimitiedon rikastamiseen?”. Työn ratkaisun myötä syntyneen datan analysoinnin avulla pystyttiin toteamaan, että Suomen ajallinen paikkaontologia soveltuu rikastamiseen hyvin, sillä löydettyistä paikannimistä reilusti yli puolet (arkistonmuodostajista 82,5 %, arkistoyksiköistä 68,2 %) kyettiin linkittämään SAPOsta löydettyihin kohteisiin. Samalla kuitenkin voidaan todeta samojen lukujen perusteella, että SAPO ei yksistään riitä paikkatiedon lähteeksi.

7.1. Yhteenveto

Työssä esiteltiin asiakkaiden tarpeita kohdistuen paikkatietoon kulttuuriaineistoissa ja arkistojen aineistoissa. Nykytilanteessa paikkatietoja kuitenkin on tuotettu ja hallinnoitu varsin vaihtelevasti, eivätkä niiden tuottaminen perustu valmiisiin sanastoihin tai vastaaviin. Tämä taas saattaa aiheuttaa haasteita aineistojen löytämisessä, kun hakuehtoina käytetään paikannimiä tai paikkoja.

Tavoitetilassa arkistolaitos on ottanut käyttöönsä uuden arkistoaineistojen kuvailu- ja luettelointitietojen hallintaan tarkoitetun järjestelmän. Järjestelmässä hallinnoitu metatieto saadaan välitettyä kirjastojen, arkistojen ja museoiden yhteiseen haku- ja palveluportaaliiin; Finnaan. Finna mahdollistaa tavoitetilassa erilaisten kuvailussa hyödynnettyjen sanastojen ja ontologioiden hyödyntämisen hauissa ja suosituksissa loppukäyttäjälle. Jotta tämä olisi mahdollista, on arkistojen rikastettava nykyisiä metatietojaan linkittämällä näitä avoimeen linkitettyyn tietoon ja ontologioihin.

Viimeisten parin vuoden aikana kansalliset ontologiat ovat saaneet pysyvemmän roolin kulttuurisektoreiden näkökulmasta, kun niiden ylläpito on keskitetty Fintoon ja Kansalliskirjastoon. Finto ylläpitää Suomen ajallista paikkaontologiaa SAPOa, joka sisältää tiedot Suomen kunnista ja näiden ajallisista kattavuuksista kuntahistoria ajalta. Arkistolaitos on päätenyt hyödyntämään ja selvittämään SAPOn nykyisen soveltuvuuden paikkatietojen rikastamisessa tulevan kuvailu- ja luettelointitietojen hallintajärjestelmän käyttöönoton yhteydessä.

Työssä määriteltiin ratkaisu ja sen vaatimukset SAPOn hyödyntämisessä nykyisten paikkatietojen rikastamisessa. Tuloksista huomattiin, että SAPO soveltuu hyvin suurelle

osalle tuotetusta paikkatiedosta. Erityisen hyvin SAPO soveltui arkistomuodostajiin liitettyyn paikkatietoon (82,5 % paikannimistä voitiin linkittää), mutta kohtalaisen hyvin (68,2 % paikannimistä voitiin linkittää) myös arkistoyksiköille merkittyyn paikkatietoon. Tämä toimii myös vastauksena tutkimuskysymykseen, jonka pohjalta oli tarkoitus selvittää miten hyvin SAPO soveltuu arkistolaitoksen paikannimitiedon rikastamiseen. Lisäksi ratkaisu osoittautui onnistuneeksi, koska SAPO soveltui hyvin tiedon rikastamiseen.

7.2. Pohdinta

Työn keskeisimpänä huomiona voidaan pitää sitä, että SAPO soveltuu arkistolaitoksessa hyvin yleiskuvailuun. Sen sijaan sisällönkuvailuun tai aiheiden kuvailuun arkistoyksikkötasolla se ei niin hyvin sovellu. Tarkempien kuin kuntatasoisten paikkojen ja ulkomaisten paikkojen osalta tulisi löytää sopiva ontologia tai avoin linkitetty datajoukko, johon paikannimiä voitaisiin kiinnittää. Tähän on olemassa niin kansallisia kuin kansainvälisiäkin vaihtoehtoja ja näiden soveltuvuus voisi olla seuraava mahdollinen tutkimuskohde. Tällä hetkellä (28.6.2015) tiedetään, että Kansalliskirjasto on tekemässä selvitystä paikkatietolähteiden osalta kansallisesti ja kansainvälisesti, mutta selvitystä ei ole vielä julkaistu. Tämä todennäköisesti voisi toimia pohjana seuraavalle arvioinnille sekä päätöksenteolle arkistolaitoksessa.

SAPO soveltuu arkistoille ja todennäköisesti myös muille historiallisia kohteita kuvailleville organisaatioille hyvin, sillä se sisältää myös historiallista paikkatietoa. Tästä hyvä esimerkki on luvun 6 alussa annettu esimerkki, jossa SAPOsta löydettiin kohde ”Kivennapa”. Kivennapa ei nykyisin ole enää Suomen kunta, mutta se on ollut kunta aina vuoteen 1940 saakka, kuten SAPOstakin huomataan (http://dev.finto.fi/sapo/fi/page/Kivennapa_1937_01_01).

Työn tulokset palvelevat myös arkistojen asiakkaita. Koska arkistonmuodostajista, joille paikkatietoa oli merkitty vähintään 82,5 % löydettiin vastine SAPOsta ja arkistoyksiköille vastaava luku oli 68,3 %, voidaan todeta, että paikannimien avulla arkistoaineistoa etsittäessä kannattaa usein käyttää kuntatason tarkkuutta. On todennäköisempää, että paikkatieto on annettu kuntatasolla, kuin millään muulla tasolla.

Se, että työn tuloksia voidaan pitää hyvinä ja hyvänä lähtökohtana uuden AHAA-järjestelmän sekä uusien kuvailu- ja luettelointisääntöjen käyttöönotolle, ei takaa vielä

luvussa 3 kuvattua tavoitetilaa. Tulosten avulla voidaan osittain perustella, että SAPO tulisi julkaista tuotantoversiona ja sen ylläpito tulisi olla tuettua, koska sen sisältämä tieto on relevanttia ja hyödyllistä arkistolaitoksen näkökulmasta. Tulosten avulla voidaan osittain perustella myös sitä, että Finnaan tulisi rakentaa AHAA-järjestelmän käyttöönoton koittaessa hakutoiminnallisuuksia, joissa SAPOa käytettäisiin hakujen tai selaamisen tukena tai apuvälineenä. Nämä ovat kuitenkin asioita, joita arkistolaitos ei yksin voi päättää vaan työtä ohjaavat rakenteet ovat ennalta asetetut. Arkistolaitoksen tulisi kuitenkin työn tuloksiin pohjautuen pystyä perustelemaan tarvittavia muutoksia seuraavien vuosien aikana KDK:n ja Finton puitteissa. Uuden toimintatavan muodostaminen mahdollistaisi myös tämän työn ratkaisun hyödyllisyyden arvioinnin. Tämä voitaisiin suorittaa esimerkiksi tekemällä uusi asiakastytyväisyyskysely.

8. Lähdeluettelo

Ahmad, M. N. & Colomb, R. M., 2007. *Managing Ontologies: A Comparative Study of Ontology Servers*. Ballarat, Victoria, Australia, Australian Computer Society Inc..

Antoniou, G. & van Harmelen, F., 2004. *A Semantic Web Primer*. London: The MIT press.

Arkistolaitos, 1997. *Arkistojen kuvailu- ja luettelointisäännöt 1997*. [Online]
Available at: <http://www.arkisto.fi/se/arkistojen-kuvailu--ja-luettelointisaaennoet>
[Haettu 6 Huhtikuu 2015].

Arkistolaitos, 2012. *AHAA-hanke*. [Online]
Available at: <http://www.arkisto.fi/sv/tjanster/arkistoyhteistyoe-2/arkistojen-hakemistopalvelu-2>
[Haettu 8 Maaliskuu 2015].

Arkistolaitos, 2014. *Asiakastytyväisyyskyselyn tulokset 2014*, s.l.: s.n.

Arkistolaitos, ei pvm *Arkistolaitoksen strategia 2015*, Helsinki: Arkistolaitos.

Arkistolaitos, ei pvm *Asiakirjahallinnan sanasto*. [Online]
Available at:
http://wiki.narc.fi/arkistowiki/index.php/Luokka:Asiakirjahallinnan_sanasto
[Haettu 8 3 2015].

Berners-Lee, T., 1998. *Uniform Resource Identifiers (URI): Generic Syntax*. s.l.:Network Working Group.

Berners-Lee, T., 2006. *Linked Data*. [Online]
Available at: <http://www.w3.org/DesignIssues/LinkedData.html>
[Haettu 3 Toukokuu 2015].

Berners-Lee, T., Hendler, J. & Lassila, O., 2001. The Semantic Web. *Scientific American*, Issue May, pp. 29-37.

Bohn, D., 2012. Google Now: behind the predictive future of search. *The Verge*, 29 Lokakuu.

Duff, W. M. ym., 2006. Archivists' views of user-based evaluation: Benefits, barriers and requirements. *The American Archivist*, 69(1).

Farber, D., 2013. Microsoft's Bing seeks enlightenment with Satori. *CNET*, 30 Heinäkuu.

Finna, 2015. *Finna työskentelysivusto*. [Online]
Available at: <https://www.kiwi.fi/display/finna/Finna>
[Haettu 14 Huhtikuu 2015].

Finto, 2014. *SAPo-ontologia (kehityspuoli)*. [Online]
Available at: <http://dev.finto.fi/sapo/fi/>
[Haettu 8 3 2015].

Finto, 2015. *Finto-palvelu*. [Online]
Available at: <http://finto.fi/fi/>
[Haettu 8 Maaliskuu 2015].

Finto, ei pvm *KOKO-ontologia*. [Online]
Available at: <http://finto.fi/koko/fi/>
[Haettu 8 3 2015].

Finto, ei pvm *YSO-ontologia*. [Online]
Available at: <http://finto.fi/ysa/fi/>
[Haettu 8 3 2015].

Fu, G., Jones, C. B. & Abdelmoty, A. I., 2005. Ontology-based Spatial Query Expansion in Information Retrieval. Teoksessa: *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE*. s.l.:Springer Berlin Heidelberg, pp. 1466-1482.

Gruber, T. R., 1993. A translation approach to portable ontology specifications. *Knowledge acquisition*, Issue 5, pp. 199-220.

Hupaniittu, O., 2011. *Tutkijoiden ääni ja sähköiset aineistot*, Helsinki: Svenska literatursällskapet i Finland.

Hübner, S., Spittel, R., Visser, U. & Vögele, T. J., 2007. Ontology-Based search for interactive digital maps. *Intelligent Systems, IEEE*, 19(3), pp. 80-86.

ICA, 2011. *ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, 2nd Edition*. [Online]
Available at: <http://www.ica.org/10203/standards/isaar-cpf-international-standard-archival-authority-record-for-corporate-bodies-persons-and-families-2nd-edition.html>
[Haettu 3 7 2015].

ICA, 2011. *ISAD(G): General International Standard Archival Description - Second edition*. [Online]
Available at: <http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html>
[Haettu 7 3 2015].

ICA, ei pvm *Discover ICA*. [Online]
Available at: <http://www.ica.org/104/presentation-of-the-ica/discover-ica.html>
[Haettu 8 3 2015].

IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009.
Functional requirements for bibliographic records, s.l.: s.n.

Internet Live Stats, 2015. *Internet Live Stats*. [Online]
Available at: <http://www.internetlivestats.com/internet-users/>
[Haettu 16 Toukokuu 2015].

Janowicz, K. & Oberle, D., 2009. How Ontologies Benefit Enterprise Applications.
Semantic web journal, pp. 1-5.

Kansallinen digitaalinen kirjasto, 2012. *KDK kokonaisarkkitehtuuri*. [Online]
Available at: <http://www.kdk.fi/images/tiedostot/KDK-kokonaisarkkitehtuuri2.pdf>
[Haettu 14 Huhtikuu 2015].

Kansallinen digitaalinen kirjasto, 2014. *KDK standardisalkku*. [Online]
Available at: <http://www.kdk.fi/fi/kokonaisarkkitehtuuri/16-suomi/kokonaisarkkitehtuuri/85-standardisalkku>
[Haettu 14 Huhtikuu 2015].

Kansallinen digitaalinen kirjasto, 2015. *Kansallinen digitaalinen kirjasto*. [Online]
Available at: <http://www.kdk.fi/fi/tietoa-hankkeesta>
[Haettu 13 Huhtikuu 2015].

Kansallinen digitaalinen kirjasto, 2015. *KDK kokonaisarkkitehtuuri*. [Online]
Available at: <http://www.kdk.fi/fi/kokonaisarkkitehtuuri>
[Haettu 14 Huhtikuu 2015].

Kansallisarkisto, 2013. *AHAA tarvekartoitus ja vaatimusmäärittely*. Helsinki: s.n.

Kansalliskirjasto, 2015. *Finna ja Kansallinen digitaalinen kirjasto*. [Online]
Available at: <http://www.kansalliskirjasto.fi/kirjastoala/asiakasliittyma.html>
[Haettu 14 Huhtikuu 2015].

Kasanen, E., Lukka, K. & Siitonen, A., 1993. The Constructive Approach in Management Accounting Research. *Journal of Management Accounting Research*, Osa/vuosikerta Vol 5., pp. 241-264.

KDK, 2009. *Kansallinen digitaalinen kirjasto*. [Online]
Available at:
http://www.kdk.fi/images/stories/KDK_asiakasliittym_toiminnallisuus_24.6.2009.pdf
[Haettu 29 Maaliskuu 2015].

Kilkki, J. & Henttonen, P., 2015. *Arkistokuvailun kansallinen käsitelmä*. [Online]
Available at:
http://www.arkisto.fi/uploads/Palvelut/Arkistoyhteisty%C3%B6/Arkistosektorin%20yhteisty%C3%B6verkosto/Arkistokuvailun_kansallinen_kasitemalli_02.pdf
[Haettu 29 3 2015].

Laukkonen, A., 2015. *am_sapo.xlsx*. Helsinki: s.n.

Laukkonen, A., 2015. *Suunnittelija* [Haastattelu] (Huhtikuu 2015).

Lei, Y., Uren, V. & Motta, E., 2006. SemSearch: A Search Engine for the Semantic Web. Teoksessa: *Proc. 5th International Conference on Knowledge Engineering and Knowledge Management Managing Knowledge in a World of Networks*. s.l.:Springer-Verlag, pp. 238-245.

Linked Data Finland, ei pvm *Linked Data Finland*. [Online]
Available at: <http://www.ldf.fi/>
[Haettu 3 Toukokuu 2015].

Maanmittauslaitos, 2015. *Maanmittauslaitos*. [Online]

Available at: <http://www.maanmittauslaitos.fi/kartat/kartoitus/paikkatiedot>

[Haettu 18.1.2015].

Semantic Computing Research Group (SeCo), 2014. *Semantic Computing Research Group (SeCo)*. [Online]

Available at: <http://www.seco.tkk.fi/>

[Haettu 15. Huhtikuu 2015].

Summers, E. & Salo, D., 2013. Linking Things on the Web: A Pragmatic Examination of Linked Data. *CoRR*, Osa/vuosikerta abs/1302.4591.

W3C, 2015. *Semantic Web*. [Online]

Available at: <http://www.w3.org/standards/semanticweb/>

[Haettu 3. Toukokuu 2015].

Wikipedia, 2014. *Wikipedia - WGS84*. [Online]

Available at: <http://fi.wikipedia.org/wiki/WGS84>

[Haettu 17. Toukokuu 2015].

Wikipedia, 2015. *Wikipedia - Ontology (information science)*. [Online]

Available at: http://en.wikipedia.org/wiki/Ontology_%28information_science%29

[Haettu 31. Toukokuu 2015].